

5

**INDUCING CELLULAR IMMUNE RESPONSES TO HEPATITIS C VIRUS USING
PEPTIDE AND NUCLEIC ACID COMPOSITIONS****CROSS-REFERENCES TO RELATED APPLICATIONS**

This application is a Continuation-In-Part ("CIP") of U.S.S.N. 09/189,702 filed
10 11/10/98, which is a CIP of U.S.S.N. 08/205,713 filed 3/4/94, which is a CIP of 08/159,184
filed 11/29/93 and now abandoned, which is a CIP of 08/073,205 filed 6/4/93 and now
abandoned, which is a CIP of 08/027,146 filed 3/5/93 and now abandoned. The present
application is also related to U.S.S.N. 09/226,775, which is a CIP of U.S.S.N. 08/815,396,
which claims the benefit of U.S.S.N. 60/013,113, now abandoned. Furthermore, the present
15 application is related to U.S.S.N. 09/017,735, which is a CIP of abandoned U.S.S.N.
08/589,108; U.S.S.N. 08/753,622, U.S.S.N. 08/822,382, abandoned U.S.S.N. 60/013,980,
U.S.S.N. 08/454,033, U.S.S.N. 09/116,424, and U.S.S.N. 08/349,177. The present
application is also related to U.S.S.N. 09/017,524, U.S.S.N. 08/821,739, abandoned U.S.S.N.
60/013,833, U.S.S.N. 08/758,409, U.S.S.N. 08/589,107, U.S.S.N. 08/451,913, U.S.S.N.
20 08/186,266, U.S.S.N. 09/116,061, and U.S.S.N. 08/347,610, which is a CIP of U.S.S.N.
08/159,339, which is a CIP of abandoned U.S.S.N. 08/103,396, which is a CIP of abandoned
U.S.S.N. 08/027,746, which is a CIP of abandoned U.S.S.N. 07/926,666. The present
application is also related to U.S.S.N. 09/017,743, U.S.S.N. 08/753,615; U.S.S.N.
08/590,298, U.S.S.N. 09/115,400, and U.S.S.N. 08/452,843, which is a CIP of U.S.S.N.
25 08/344,824, which is a CIP of abandoned U.S.S.N. 08/278,634. The present application is
also related to provisional U.S.S.N. 60/087,192 and U.S.S.N. 09/009,953, which is a CIP of
abandoned U.S.S.N. 60/036,713 and abandoned U.S.S.N. 60/037,432. In addition, the
present application is related to U.S.S.N. 09/098,584, U.S.S.N. 09/239,043, and to
Provisional U.S. Patent Application 60/117,486 filed 1/27/99. The present application is
30 also related to U.S. Patent Application entitled "Inducing Cellular Immune Responses to
Hepatitis C Virus Using Peptide and Nucleic Acid Compositions", Attorney Docket No.
018623-0013910 filed 7/8/99. All of the above applications are incorporated herein by
reference.

FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

This invention was funded, in part, by the United States government under grants with the National Institutes of Health. The U.S. government has certain rights in this invention.

INDEX

	I.	Background of the Invention
	II.	Summary of the Invention
	III.	Brief Description of the Figures
5	IV.	Detailed Description of the Invention
	A.	Definitions
	B.	Stimulation of CTL and HTL responses
	C.	Binding Affinity of Peptide Epitopes for HLA Molecules
	D.	Peptide Epitope Binding Motifs and Supermotifs
10	1.	HLA-A1 supermotif
	2.	HLA-A2 supermotif
	3.	HLA-A3 supermotif
	4.	HLA-A24 supermotif
	5.	HLA-B7 supermotif
15	6.	HLA-B27 supermotif
	7.	HLA-B44 supermotif
	8.	HLA-B58 supermotif
	9.	HLA-B62 supermotif
	10.	HLA-A1 motif
20	11.	HLA-A2.1 motif
	12.	HLA-A3 motif
	13.	HLA-A11 motif
	14.	HLA-A24 motif
	15.	HLA-DR-1-4-7 supermotif
25	16.	HLA-DR3 motifs
	E.	Enhancing Population Coverage of the Vaccine
	F.	Immune Response-Stimulating Peptide Epitope Analogs
	G.	Computer Screening of Protein Sequences from Disease-Related Antigens for Supermotif- or Motif-Containing Epitopes
30	H.	Preparation of Peptide Epitopes

- I. Assays to Detect T-Cell Responses
- J. Use of Peptide Epitopes for Evaluating Immune Responses
- K. Vaccine Compositions
 - 1. Minigene Vaccines
 - 5 2. Combinations of CTL Peptides with Helper Peptides
- L. Administration of Vaccines for Therapeutic or Prophylactic Purposes
- M. Kits
- V. Examples
- VI. Claims
- 10 VII. Abstract

I. BACKGROUND OF THE INVENTION

Hepatitis C virus (HCV) infection is a global human health problem with approximately 150,000 new reported cases each year in the U.S. alone. HCV is a single stranded RNA virus, and is the etiological agent identified in most cases of non-A, non-B post-transfusion and post-transplant hepatitis, and is a common cause of acute sporadic hepatitis (Choo *et al.*, *Science* 244:359, 1989; Kuo *et al.*, *Science* 244:362, 1989; and Alter *et al.*, in: *Current Perspective in Hepatology*, p. 83, 1989). It is estimated that more than 50% of patients infected with HCV become chronically infected and, of those, 20% develop cirrhosis of the liver within 20 years (Davis *et al.*, *New Engl. J. Med.* 321:1501, 1989; Alter *et al.*, in: *Current Perspective in Hepatology*, p. 83, 1989; Alter *et al.*, *New Engl. J. Med.* 327:1899, 1992; and Dienstag, J. L. *Gastroenterology* 85:430, 1983). Moreover, the only therapy available for treatment of HCV infection is interferon- α . Most patients are unresponsive, however, and among the responders, there is a high recurrence rate within 6-12 months of cessation of treatment (Liang *et al.*, *J. Med. Virol.* 40:69, 1993). Ribavirin, a guanosine analog with a broad spectrum activity against many RNA and DNA viruses, has been shown in clinical trials to be effective against chronic HCV infection when used in combination with interferon- α (*see, e.g.*, Poynard *et al.*, *Lancet* 352:1426-1432, 1998; Reichard *et al.*, *Lancet* 351:83-87, 1998). However, the response rate is still well below 50%.

Virus-specific, human leukocyte antigen (HLA) class I-restricted cytotoxic T lymphocytes (CTL) are known to play a major role in the prevention and clearance of virus infections *in vivo* (Oldstone *et al.*, *Nature* 321:239, 1989; Jamieson *et al.*, *J. Virol.* 61:3930, 1987; Yap *et al.*, *Nature* 273:238, 1978; Lukacher *et al.*, *J. Exp. Med.* 160:814, 1994; McMichael *et al.*, *N. Engl. J. Med.* 309:13, 1983; Sethi *et al.*, *J. Gen. Virol.* 64:443, 1983; Watari *et al.*, *J. Exp. Med.* 165:459, 1987; Yasukawa *et al.*, *J. Immunol.* 143:2051, 1989; Tigges *et al.*, *J. Virol.* 66:1622, 1993; Reddenhase *et al.*, *J. Virol.* 55:263, 1985; Quinnan *et al.*, *N. Engl. J. Med.* 307:6, 1982). HLA class I molecules are expressed on the surface of almost all nucleated cells. Following intracellular processing of antigens, epitopes from the antigens are presented as a complex with the HLA class I molecules on the surface of such cells. CTL recognize the peptide-HLA class I complex, which then results in the destruction of the cell bearing the HLA-peptide complex directly by the CTL and or via the activation of non-destructive mechanisms *e.g.*, the production of interferon, that inhibit viral replication.

In view of the heterogeneous immune response observed with HCV infection, induction of a multi-specific cellular immune response directed simultaneously against multiple HCV epitopes appears to be important for the development of an efficacious vaccine against HCV. There is a need, however, to establish vaccine embodiments that elicit immune responses that correspond to responses seen in patients that clear HCV infection.

The information provided in this section is intended to disclose the presently understood state of the art as of the filing date of the present application. Information is included in this section which was generated subsequent to the priority date of this application. Accordingly, information in this section is not intended, in any way, to delineate the priority date for the invention.

II. SUMMARY OF THE INVENTION

This invention applies our knowledge of the mechanisms by which antigen is recognized by T cells, for example, to develop epitope-based vaccines directed towards HCV. More specifically, this application communicates our discovery of specific epitope pharmaceutical compositions and methods of use in the prevention and treatment of HCV infection.

Upon development of appropriate technology, the use of epitope-based vaccines has several advantages over current vaccines, particularly when compared to the use of whole antigens in vaccine compositions. There is evidence that the immune response to whole antigens is directed largely toward variable regions of the antigen, allowing for immune escape due to mutations. The epitopes for inclusion in an epitope-based vaccine are selected from conserved regions of viral or tumor-associated antigens, which thereby reduces the likelihood of escape mutants. Furthermore, immunosuppressive epitopes that may be present in whole antigens can be avoided with the use of epitope-based vaccines.

An additional advantage of an epitope-based vaccine approach is the ability to combine selected epitopes (CTL and HTL), and further, to modify the composition of the epitopes, achieving, for example, enhanced immunogenicity. Accordingly, the immune response can be modulated, as appropriate, for the target disease. Similar engineering of the response is not possible with traditional approaches.

Another major benefit of epitope-based immune-stimulating vaccines is their safety. The possible pathological side effects caused by infectious agents or whole protein antigens, which might have their own intrinsic biological activity, is eliminated.

An epitope-based vaccine also provides the ability to direct and focus an immune response to multiple selected antigens from the same pathogen. Thus, patient-by-patient variability in the immune response to a particular pathogen may be alleviated by inclusion of epitopes from multiple antigens from that pathogen in a vaccine composition. A

5 "pathogen" may be an infectious agent or a tumor associated molecule.

One of the most formidable obstacles to the development of broadly efficacious epitope-based immunotherapeutics, however, has been the extreme polymorphism of HLA molecules. To date, effective non-genetically biased coverage of a population has been a task of considerable complexity; such coverage has required that epitopes be used
10 that are specific for HLA molecules corresponding to each individual HLA allele, therefore, impractically large numbers of epitopes would have to be used in order to cover ethnically diverse populations. Thus, there has existed a need for peptide epitopes that are bound by multiple HLA antigen molecules for use in epitope-based vaccines. The greater the number of HLA antigen molecules bound, the greater the breadth of
15 population coverage by the vaccine.

Furthermore, as described herein in greater detail, a need has existed to modulate peptide binding properties, for example, so that peptides that are able to bind to multiple HLA antigens do so with an affinity that will stimulate an immune response. Identification of epitopes restricted by more than one HLA allele at an affinity that
20 correlates with immunogenicity is important to provide thorough population coverage, and to allow the elicitation of responses of sufficient vigor to prevent or clear an infection in a diverse segment of the population. Such a response can also target a broad array of epitopes. The technology disclosed herein provides for such favored immune responses.

In a preferred embodiment, epitopes for inclusion in vaccine compositions of the
25 invention are selected by a process whereby protein sequences of known antigens are evaluated for the presence of motif or supermotif-bearing epitopes. Peptides corresponding to a motif- or supermotif-bearing epitope are then synthesized and tested for the ability to bind to the HLA molecule that recognizes the selected motif. Those peptides that bind at an intermediate or high affinity *i.e.*, an IC_{50} (or a K_D value) of 500
30 nM or less for HLA class I molecules or 1000 nM or less for HLA class II molecules, are further evaluated for their ability to induce a CTL or HTL response. Immunogenic peptide epitopes are selected for inclusion in vaccine compositions.

Supermotif-bearing peptides may additionally be tested for the ability to bind to multiple alleles within the HLA supertype family. Moreover, peptide epitopes may be

analogued to modify binding affinity and/or the ability to bind to multiple alleles within an HLA supertype.

The invention also includes an embodiment comprising a method for monitoring or evaluating an immune response to HCV in a patient having a known HLA-type, the method comprising incubating a T lymphocyte sample from the patient with a peptide composition comprising an HCV epitope consisting essentially of an amino acid sequence described in Tables VII to Table XX or Table XXII which binds the product of at least one HLA allele present in said patient, and detecting for the presence of a T lymphocyte that binds to the peptide. A CTL peptide epitope may, for example, comprise a tetrameric complex.

An alternative modality for defining the peptide epitopes in accordance with the invention is to recite the physical properties, such as length; primary structure; or charge, which are correlated with binding to a particular allele-specific HLA molecule or group of allele-specific HLA molecules. A further modality for defining peptide epitopes is to recite the physical properties of an HLA binding pocket, or properties shared by several allele-specific HLA binding pockets (*e.g.* pocket configuration and charge distribution) and reciting that the peptide epitope fits and binds to said pocket or pockets.

As will be apparent from the discussion below, other methods and embodiments are also contemplated. Further, novel synthetic peptides produced by any of the methods described herein are also part of the invention.

III. BRIEF DESCRIPTION OF THE FIGURES

Figure 1: Figure 1 provides a graph of total frequency of genotypes as a function of the number of HCV candidate epitopes bound by HLA-A and B molecules, in an average population.

Figure 2: Figure 2 illustrates the position of peptide epitopes in an experimental model minigene construct.

IV. DETAILED DESCRIPTION OF THE INVENTION

The peptide epitopes and corresponding nucleic acid compositions of the present invention are useful for stimulating an immune response to HCV by stimulating the production of CTL or HTL responses. The peptide epitopes, which are derived directly or indirectly from native HCV amino acid sequences, are able to bind to HLA molecules and stimulate an immune response to HCV. The complete polyprotein sequence from HCV

and its variants can be obtained from Genbank. Peptide epitopes and analogs thereof can also be readily determined from sequence information that may subsequently be discovered for heretofore unknown variants of HCV, as will be clear from the disclosure provided below.

5 The peptide epitopes of the invention have been identified in a number of ways, as will be discussed below. Also discussed in greater detail is that analog peptides have been derived and the binding activity for HLA molecules modulated by modifying specific amino acid residues to create peptide analogs exhibiting altered immunogenicity. Further, the present invention provides compositions and combinations of compositions
10 that enable epitope-based vaccines that are capable of interacting with HLA molecules encoded by various genetic alleles to provide broader population coverage than prior vaccines.

IV.A. Definitions

15 The invention can be better understood with reference to the following definitions, which are listed alphabetically:

 A "computer" or "computer system" generally includes: a processor; at least one information storage/retrieval apparatus such as, for example, a hard drive, a disk drive or a tape drive; at least one input apparatus such as, for example, a keyboard, a mouse, a
20 touch screen, or a microphone; and display structure. Additionally, the computer may include a communication channel in communication with a network. Such a computer may include more or less than what is listed above.

 "Cross-reactive binding" indicates that a peptide is bound by more than one HLA molecule; a synonym is degenerate binding.

25 A "cryptic epitope" elicits a response by immunization with an isolated peptide, but the response is not cross-reactive *in vitro* when intact whole protein which comprises the epitope is used as an antigen.

 A "dominant epitope" is an epitope that induces an immune response upon immunization with a whole native antigen (see, *e.g.*, Sercarz, *et al.*, *Annu. Rev. Immunol.*
30 11:729-766, 1993). Such a response is cross-reactive *in vitro* with an isolated peptide epitope.

 With regard to a particular amino acid sequence, an "epitope" is a set of amino acid residues which is involved in recognition by a particular immunoglobulin, or in the context of T cells, those residues necessary for recognition by T cell receptor proteins

and/or Major Histocompatibility Complex (MHC) receptors. In an immune system setting, *in vivo* or *in vitro*, an epitope is the collective features of a molecule, such as primary, secondary and tertiary peptide structure, and charge, that together form a site recognized by an immunoglobulin, T cell receptor or HLA molecule. Throughout this disclosure epitope and peptide are often used interchangeably. It is to be appreciated, however, that isolated or purified protein or peptide molecules larger than and comprising an epitope of the invention are still within the bounds of the invention.

"Human Leukocyte Antigen" or "HLA" is a human class I or class II Major Histocompatibility Complex (MHC) protein (*see, e.g.,* Stites, *et al.*, IMMUNOLOGY, 8th ED., Lange Publishing, Los Altos, CA (1994)).

An "HLA supertype or family", as used herein, describes sets of HLA molecules grouped on the basis of shared peptide-binding specificities. HLA class I molecules that share somewhat similar binding affinity for peptides bearing certain amino acid motifs are grouped into HLA superotypes. The terms HLA superfamily, HLA supertype family, HLA family, and HLA xx-like supertype molecules (where xx denotes a particular HLA type), are synonyms.

Throughout this disclosure, results are expressed in terms of "IC₅₀'s." IC₅₀ is the concentration of peptide in a binding assay at which 50% inhibition of binding of a reference peptide is observed. Given the conditions in which the assays are run (*i.e.*, limiting HLA proteins and labeled peptide concentrations), these values approximate K_D values. It should be noted that IC₅₀ values can change, often dramatically, if the assay conditions are varied, and depending on the particular reagents used (*e.g.*, HLA preparation, *etc.*). For example, excessive concentrations of HLA molecules will increase the apparent measured IC₅₀ of a given ligand.

Assays for determining binding are described in detail, *e.g.*, in PCT publications WO 94/20127 and WO 94/03205. Alternatively, binding is expressed relative to a reference peptide. As a particular assay becomes more, or less, sensitive, the IC₅₀'s of the peptides tested may change somewhat. However, the binding relative to the reference peptide will not significantly change. For example, in an assay run under conditions such that the IC₅₀ of the reference peptide increases 10-fold, the IC₅₀ values of the test peptides will also shift approximately 10-fold. Therefore, to avoid ambiguities, the assessment of whether a peptide is a good, intermediate, weak, or negative binder is generally based on its IC₅₀, relative to the IC₅₀ of a standard peptide.

Binding may also be determined using other assay systems including those using:
 live cells (e.g., Ceppellini *et al.*, *Nature* 339:392, 1989; Christnick *et al.*, *Nature* 352:67,
 1991; Busch *et al.*, *Int. Immunol.* 2:443, 1990; Hill *et al.*, *J. Immunol.* 147:189, 1991; del
 Guercio *et al.*, *J. Immunol.* 154:685, 1995), cell free systems using detergent lysates (e.g.,
 5 Cerundolo *et al.*, *J. Immunol.* 21:2069, 1991), immobilized purified MHC (e.g., Hill *et*
al., *J. Immunol.* 152, 2890, 1994; Marshall *et al.*, *J. Immunol.* 152:4946, 1994), ELISA
 systems (e.g., Reay *et al.*, *EMBO J.* 11:2829, 1992), surface plasmon resonance (e.g.,
 Khilko *et al.*, *J. Biol. Chem.* 268:15425, 1993); high flux soluble phase assays (Hammer
et al., *J. Exp. Med.* 180:2353, 1994), and measurement of class I MHC stabilization or
 10 assembly (e.g., Ljunggren *et al.*, *Nature* 346:476, 1990; Schumacher *et al.*, *Cell* 62:563,
 1990; Townsend *et al.*, *Cell* 62:285, 1990; Parker *et al.*, *J. Immunol.* 149:1896, 1992).

As used herein, "high affinity" with respect to HLA class I molecules is defined as
 binding with an IC_{50} , or K_D value, of 50 nM or less; "intermediate affinity" is binding
 with an IC_{50} or K_D value of between about 50 and about 500 nM. "High affinity" with
 15 respect to binding to HLA class II molecules is defined as binding with an IC_{50} or K_D
 value of 100 nM or less; "intermediate affinity" is binding with an IC_{50} or K_D value of
 between about 100 and about 1000 nM.

The terms "identical" or percent "identity," in the context of two or more peptide
 sequences, refer to two or more sequences or subsequences that are the same or have a
 20 specified percentage of amino acid residues that are the same, when compared and
 aligned for maximum correspondence over a comparison window, as measured using a
 sequence comparison algorithm or by manual alignment and visual inspection.

An "immunogenic peptide" or "peptide epitope" is a peptide that comprises an
 allele-specific motif or supermotif such that the peptide will bind an HLA molecule and
 25 induce a CTL and/or HTL response. Thus, immunogenic peptides of the invention are
 capable of binding to an appropriate HLA molecule and thereafter inducing a cytotoxic T
 cell response, or a helper T cell response, to the antigen from which the immunogenic
 peptide is derived.

The phrases "isolated" or "biologically pure" refer to material which is
 30 substantially or essentially free from components which normally accompany the material
 as it is found in its native state. Thus, isolated peptides in accordance with the invention
 preferably do not contain materials normally associated with the peptides in their *in situ*
 environment.

“Major Histocompatibility Complex” or “MHC” is a cluster of genes that plays a role in control of the cellular interactions responsible for physiologic immune responses. In humans, the MHC complex is also known as the HLA complex. For a detailed description of the MHC and HLA complexes, see, Paul, FUNDAMENTAL IMMUNOLOGY, 3RD ED., Raven Press, New York, 1993.

The term "motif" refers to the pattern of residues in a peptide of defined length, usually a peptide of from about 8 to about 13 amino acids for a class I HLA motif and from about 6 to about 25 amino acids for a class II HLA motif, which is recognized by a particular HLA molecule. Peptide motifs are typically different for each protein encoded by each human HLA allele and differ in the pattern of the primary and secondary anchor residues.

A "negative binding residue" is an amino acid which, if present at certain positions (typically not primary anchor positions) in a peptide epitope, results in decreased binding affinity of the peptide for the peptide's corresponding HLA molecule.

The term "peptide" is used interchangeably with "oligopeptide" in the present specification to designate a series of residues, typically L-amino acids, connected one to the other, typically by peptide bonds between the α -amino and carboxyl groups of adjacent amino acids. The preferred CTL-inducing peptides of the invention are 13 residues or less in length and usually consist of between about 8 and about 11 residues, preferably 9 or 10 residues. The preferred HTL-inducing oligopeptides are less than about 50 residues in length and usually consist of between about 6 and about 30 residues, more usually between about 12 and 25, and often between about 15 and 20 residues.

“Pharmaceutically acceptable” refers to a non-toxic, inert, and physiologically compatible composition.

A "primary anchor residue" is an amino acid at a specific position along a peptide sequence which is understood to provide a contact point between the immunogenic peptide and the HLA molecule. One to three, usually two, primary anchor residues within a peptide of defined length generally defines a “motif” for an immunogenic peptide. These residues are understood to fit in close contact with peptide binding grooves of an HLA molecule, with their side chains buried in specific pockets of the binding grooves themselves. In one embodiment, the primary anchor residues are located at position 2 (from the amino terminal position) and at the carboxyl terminal position of a 9-residue peptide epitope in accordance with the invention. The primary anchor positions for each motif and supermotif are set forth in Table 1. For example, analog peptides can

be created by altering the presence or absence of particular residues in these primary anchor positions. Such analogs are used to modulate the binding affinity of a peptide comprising a particular motif or supermotif.

5 "Promiscuous recognition" is where a distinct peptide is recognized by the same T cell clone in the context of various HLA molecules. Promiscuous recognition or binding is synonymous with cross-reactive binding.

A "protective immune response" or "therapeutic immune response" refers to a CTL and/or an HTL response to an antigen derived from an infectious agent or a tumor antigen, which prevents or at least partially arrests disease symptoms or progression. The
10 immune response may also include an antibody response which has been facilitated by the stimulation of helper T cells.

The term "residue" refers to an amino acid or amino acid mimetic incorporated into an oligopeptide by an amide bond or amide bond mimetic.

A "secondary anchor residue" is an amino acid at a position other than a primary
15 anchor position in a peptide which may influence peptide binding. A secondary anchor residue occurs at a significantly higher frequency amongst bound peptides than would be expected by random distribution of amino acids at one position. The secondary anchor residues are said to occur at "secondary anchor positions." A secondary anchor residue can be identified as a residue which is present at a higher frequency among high or
20 intermediate affinity binding peptides, or a residue otherwise associated with high or intermediate affinity binding. For example, analog peptides can be created by altering the presence or absence of particular residues in these secondary anchor positions. Such analogs are used to finely modulate the binding affinity of a peptide comprising a particular motif or supermotif.

25 A "subdominant epitope" is an epitope which evokes little or no response upon immunization with whole antigens which comprise the epitope, but for which a response can be obtained by immunization with an isolated peptide, and this response (unlike the case of cryptic epitopes) is detected when whole protein is used to recall the response *in vitro* or *in vivo*.

30 A "supermotif" is a peptide binding specificity shared by HLA molecules encoded by two or more HLA alleles. Preferably, a supermotif-bearing peptide is recognized with high or intermediate affinity (as defined herein) by two or more HLA antigens.

"Synthetic peptide" refers to a peptide that is not naturally occurring, but is man-made using such methods as chemical synthesis or recombinant DNA technology.

The nomenclature used to describe peptide compounds follows the conventional practice wherein the amino group is presented to the left (the N-terminus) and the carboxyl group to the right (the C-terminus) of each amino acid residue. When amino acid residue positions are referred to in a peptide epitope they are numbered in an amino to carboxyl direction with position one being the position closest to the amino terminal end of the epitope, or the peptide or protein of which it may be a part. In the formulae representing selected specific embodiments of the present invention, the amino- and carboxyl-terminal groups, although not specifically shown, are in the form they would assume at physiologic pH values, unless otherwise specified. In the amino acid structure formulae, each residue is generally represented by standard three letter or single letter designations. The L-form of an amino acid residue is represented by a capital single letter or a capital first letter of a three-letter symbol, and the D-form for those amino acids having D-forms is represented by a lower case single letter or a lower case three letter symbol. Glycine has no asymmetric carbon atom and is simply referred to as "Gly" or G.

15 Symbols for the amino acids are shown below.

Single Letter Symbol	Three Letter Symbol	Amino Acids
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic Acid
E	Glu	Glutamic Acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

IV.B. Stimulation of CTL and HTL responses

5 The mechanism by which T cells recognize antigens has been delineated during the past ten years. Based on our understanding of the immune system we have developed efficacious peptide epitope vaccine compositions that can induce a therapeutic or prophylactic immune response to HCV in a broad population. For an understanding of the value and efficacy of the claimed compositions, a brief review of immunology-related technology is provided.

10 A complex of an HLA molecule and a peptidic antigen acts as the ligand recognized by HLA-restricted T cells (Buus, S. *et al.*, *Cell* 47:1071, 1986; Babbitt, B. P. *et al.*, *Nature* 317:359, 1985; Townsend, A. and Bodmer, H., *Annu. Rev. Immunol.* 7:601,

1989; Germain, R. N., *Annu. Rev. Immunol.* 11:403, 1993). Through the study of single amino acid substituted antigen analogs and the sequencing of endogenously bound, naturally processed peptides, critical residues that correspond to motifs required for specific binding to HLA antigen molecules have been identified and are described herein and are set forth in Tables I, II, and III (see also, e.g., Southwood, *et al.*, *J. Immunol.* 160:3363, 1998; Rammensee, *et al.*, *Immunogenetics* 41:178, 1995; Rammensee *et al.*, SYFPEITHI, access via web at : <http://134.2.96.221/scripts.hlaserver.dll/home.htm>; Sette, A. and Sidney, J. *Curr. Opin. Immunol.* 10:478, 1998; Engelhard, V. H., *Curr. Opin. Immunol.* 6:13, 1994; Sette, A. and Grey, H. M., *Curr. Opin. Immunol.* 4:79, 1992; Sinigaglia, F. and Hammer, J. *Curr. Biol.* 6:52, 1994; Ruppert *et al.*, *Cell* 74:929-937, 1993; Kondo *et al.*, *J. Immunol.* 155:4307-4312, 1995; Sidney *et al.*, *J. Immunol.* 157:3480-3490, 1996; Sidney *et al.*, *Human Immunol.* 45:79-93, 1996; Sette, A. and Sidney, J. *Immunogenetics*, in press, 1999).

Furthermore, x-ray crystallographic analysis of HLA-peptide complexes has revealed pockets within the peptide binding cleft of HLA molecules which accommodate, in an allele-specific mode, residues borne by peptide ligands; these residues in turn determine the HLA binding capacity of the peptides in which they are present. (See, e.g., Madden, D.R. *Annu. Rev. Immunol.* 13:587, 1995; Smith, *et al.*, *Immunity* 4:203, 1996; Fremont *et al.*, *Immunity* 8:305, 1998; Stern *et al.*, *Structure* 2:245, 1994; Jones, E.Y. *Curr. Opin. Immunol.* 9:75, 1997; Brown, J. H. *et al.*, *Nature* 364:33, 1993; Guo, H. C. *et al.*, *Proc. Natl. Acad. Sci. USA* 90:8053, 1993; Guo, H. C. *et al.*, *Nature* 360:364, 1992; Silver, M. L. *et al.*, *Nature* 360:367, 1992; Matsumura, M. *et al.*, *Science* 257:927, 1992; Madden *et al.*, *Cell* 70:1035, 1992; Fremont, D. H. *et al.*, *Science* 257:919, 1992; Saper, M. A., Bjorkman, P. J. and Wiley, D. C., *J. Mol. Biol.* 219:277, 1991.)

Accordingly, the definition of class I and class II allele-specific HLA binding motifs, or class I or class II supermotifs allows identification of regions within a protein that have the potential of binding particular HLA antigen(s).

The present inventors have found that the correlation of binding affinity with immunogenicity, which is disclosed herein, is an important factor to be considered when evaluating candidate peptides. Thus, by a combination of motif searches and HLA-peptide binding assays, candidates for epitope-based vaccines have been identified. After determining their binding affinity, additional confirmatory work can be performed to select, amongst these vaccine candidates, epitopes with preferred characteristics in terms of population coverage, antigenicity, and immunogenicity.

Various strategies can be utilized to evaluate immunogenicity, including:

- 1) Evaluation of primary T cell cultures from normal individuals (*see, e.g.,* Wentworth, P. A. *et al.*, *Mol. Immunol.* 32:603, 1995; Celis, E. *et al.*, *Proc. Natl. Acad. Sci. USA* 91:2105, 1994; Tsai, V. *et al.*, *J. Immunol.* 158:1796, 1997; Kawashima, I. *et al.*, *Human Immunol.* 59:1, 1998); This procedure involves the stimulation of peripheral blood lymphocytes (PBL) from normal subjects with a test peptide in the presence of antigen presenting cells *in vitro* over a period of several weeks. T cells specific for the peptide become activated during this time and are detected using, *e.g.,* a ^{51}Cr -release assay involving peptide sensitized target cells.
- 2) Immunization of HLA transgenic mice (*see, e.g.,* Wentworth, P. A. *et al.*, *J. Immunol.* 26:97, 1996; Wentworth, P. A. *et al.*, *Int. Immunol.* 8:651, 1996; Alexander, J. *et al.*, *J. Immunol.* 159:4753, 1997); In this method, peptides in incomplete Freund's adjuvant are administered subcutaneously to HLA transgenic mice. Several weeks following immunization, splenocytes are removed and cultured *in vitro* in the presence of test peptide for approximately one week. Peptide-specific T cells are detected using, *e.g.,* a ^{51}Cr -release assay involving peptide sensitized target cells and target cells expressing endogenously generated antigen.
- 3) Demonstration of recall T cell responses from immune individuals who have effectively been vaccinated, recovered from infection, and/or from chronically infected patients (*see, e.g.,* Rehermann, B. *et al.*, *J. Exp. Med.* 181:1047, 1995; Doolan, D. L. *et al.*, *Immunity* 7:97, 1997; Bertoni, R. *et al.*, *J. Clin. Invest.* 100:503, 1997; Threlkeld, S. C. *et al.*, *J. Immunol.* 159:1648, 1997; Diepolder, H. M. *et al.*, *J. Virol.* 71:6011, 1997). In applying this strategy, recall responses are detected by culturing PBL from subjects that have been naturally exposed to the antigen, for instance through infection, and thus have generated an immune response "naturally", or from patients who were vaccinated against the infection. PBL from subjects are cultured *in vitro* for 1-2 weeks in the presence of test peptide plus antigen presenting cells (APC) to allow activation of "memory" T cells, as compared to "naive" T cells. At the end of the culture period, T cell activity is detected using assays for T cell activity including ^{51}Cr release involving peptide-sensitized targets, T cell proliferation, or lymphokine release.

The following describes the peptide epitopes and corresponding nucleic acids of the invention.

IV.C. Binding Affinity of Peptide Epitopes for HLA Molecules

As indicated herein, the large degree of HLA polymorphism is an important factor to be taken into account with the epitope-based approach to vaccine development. To address this factor, epitope selection encompassing identification of peptides capable of binding at high or intermediate affinity to multiple HLA molecules is preferably utilized, most preferably these epitopes bind at high or intermediate affinity to two or more allele specific HLA molecules.

CTL-inducing peptides of interest for vaccine compositions preferably include those that have an IC_{50} or binding affinity value for class I HLA molecules of 500 nM or better (*i.e.*, the value is ≤ 500 nM). HTL-inducing peptides preferably include those that have an IC_{50} or binding affinity value for class II HLA molecules of 1000 nM or better, (*i.e.*, the value is $\leq 1,000$ nM). For example, peptide binding is assessed by testing the capacity of a candidate peptide to bind to a purified HLA molecule *in vitro*. Peptides exhibiting high or intermediate affinity are then considered for further analysis. Selected peptides are tested on other members of the supertype family. In preferred embodiments, peptides that exhibit cross-reactive binding are then used in vaccines or in cellular screening analyses.

As disclosed herein, higher HLA binding affinity is correlated with greater immunogenicity. Greater immunogenicity can be manifested in several different ways. Immunogenicity corresponds to whether an immune response is elicited at all, and to the vigor of any particular response, as well as to the extent of a population in which a response is elicited. For example, a peptide might elicit an immune response in a diverse array of the population, yet in no instance produce a vigorous response. In accordance with these principles, close to 90% of high binding peptides have been found to be immunogenic, as contrasted with about 50% of the peptides which bind with intermediate affinity. Moreover, higher binding affinity peptides leads to more vigorous immunogenic responses. As a result, less peptide is required to elicit a similar biological effect if a high affinity binding peptide is used. Thus, in preferred embodiments of the invention, high affinity binding epitopes are particularly useful.

The relationship between binding affinity for HLA class I molecules and immunogenicity of discrete peptide epitopes on bound antigens has been determined for the first time in the art by the present inventors. The correlation between binding affinity and immunogenicity was analyzed in two different experimental approaches (*see, e.g.*,

Sette, *et al.*, *J. Immunol.* 153:5586-5592, 1994). In the first approach, the immunogenicity of potential epitopes ranging in HLA binding affinity over a 10,000-fold range was analyzed in HLA-A*0201 transgenic mice. In the second approach, the antigenicity of approximately 100 different hepatitis B virus (HBV)-derived potential epitopes, all carrying A*0201 binding motifs, was assessed by using PBL from acute hepatitis patients. Pursuant to these approaches, it was determined that an affinity threshold value of approximately 500 nM (preferably 50 nM or less) determines the capacity of a peptide epitope to elicit a CTL response. These data are true for class I binding affinity measurements for naturally processed peptides and for synthesized T cell epitopes. These data also indicate the important role of determinant selection in the shaping of T cell responses (*see, e.g.*, Schaeffer *et al. Proc. Natl. Acad. Sci. USA* 86:4649-4653, 1989).

An affinity threshold associated with immunogenicity in the context of HLA class II DR molecules has also been delineated (*see, e.g.*, Southwood *et al. J. Immunology* 160:3363-3373, 1998, and U.S.S.N. 60/087192 filed 5/29/98). In order to define a biologically significant threshold of DR binding affinity, a database of the binding affinities of 32 DR-restricted epitopes for their restricting element (*i.e.*, the HLA molecule that binds the motif) was compiled. In approximately half of the cases (15 of 32 epitopes), DR restriction was associated with high binding affinities, *i.e.* binding affinity values of 100 nM or less. In the other half of the cases (16 of 32), DR restriction was associated with intermediate affinity (binding affinity values in the 100-1000 nM range). In only one of 32 cases was DR restriction associated with an IC₅₀ of 1000 nM or greater. Thus, 1000 nM can be defined as an affinity threshold associated with immunogenicity in the context of DR molecules.

The binding affinity of peptides for HLA molecules can be determined as described in Example 1, below.

IV.D. Peptide Epitope Binding Motifs and Supermotifs

In the past few years evidence has accumulated to demonstrate that a large fraction of HLA class I and class II molecules can be classified into a relatively few supertypes, each characterized by largely overlapping peptide binding repertoires, and consensus structures of the main peptide binding pockets.

For HLA molecule pocket analyses, the residues comprising the B and F pockets of HLA class I molecules as described in crystallographic studies were analyzed (*see,*

e.g., Guo, H. C. *et al.*, *Nature* 360:364, 1992; Saper, M. A., Bjorkman, P. J. and Wiley, D. C., *J. Mol. Biol.* 219:277, 1991; Madden, D. R., Garboczi, D. N. and Wiley, D. C., *Cell* 75:693, 1993; Parham, P., Adams, E. J., and Arnett, K. L., *Immunol. Rev.* 143:141, 1995). In these analyses, residues 9, 45, 63, 66, 67, 70, and 99 were considered to make up the B pocket; and the B pocket was deemed to determine the specificity for the amino acid residue in the second position of peptide ligands. Similarly, residues 77, 80, 81, and 116 were considered to determine the specificity of the F pocket; the F pocket was deemed to determine the specificity for the C-terminal residue of a peptide ligand bound by the HLA class I molecule.

Through the study of single amino acid substituted antigen analogs and the sequencing of endogenously bound, naturally processed peptides, critical residues required for allele-specific binding to HLA molecules have been identified. The presence of these residues correlates with binding affinity for HLA molecules. The identification of motifs and/or supermotifs that correlate with high and intermediate affinity binding is an important issue with respect to the identification of immunogenic peptide epitopes for the inclusion in a vaccine. Kast *et al.* (*J. Immunol.* 152:3904-3912, 1994) have shown that motif-bearing peptides account for 90% of the epitopes that bind to allele-specific HLA class I molecules. In this study all possible peptides of 9 amino acids in length and overlapping by eight amino acids (240 peptides), which cover the entire sequence of the E6 and E7 proteins of human papillomavirus type 16, were evaluated for binding to five allele-specific HLA molecules that are expressed at high frequency among different ethnic groups. This unbiased set of peptides allowed an evaluation of the predictive value of HLA class I motifs. From the set of 240 peptides, 22 peptides were identified that bound to an allele-specific HLA molecule with high or intermediate affinity. Of these 22 peptides, 20 (*i.e.* 91%) were motif-bearing. Thus, this study demonstrates the value of motifs for the identification of peptide epitopes for inclusion in a vaccine: application of motif-based identification techniques eliminates screening of 90% of the potential epitopes in a target antigen protein sequence.

Such peptide epitopes are identified in the Tables described below.

Peptides of the present invention may also comprise epitopes that bind to MHC class II DR molecules. A greater degree of heterogeneity in both size and binding frame position of the motif, relative to the N and C termini of the peptide, exists for class II peptide ligands. This increased heterogeneity of HLA class II peptide ligands is due to the structure of the binding groove of the HLA class II molecule which, unlike its class I

counterpart, is open at both ends. Crystallographic analysis of HLA class II DRB*0101-peptide complexes showed that the major energy of binding is contributed by peptide residues complexed with complementary pockets on the DRB*0101 molecules. An important anchor residue engages the deepest hydrophobic pocket (*see, e.g.,* Madden, D.R. *Ann. Rev. Immunol.* 13:587, 1995) and is referred to as position 1 (P1). P1 may represent the N-terminal residue of a class II binding peptide epitope, but more typically is flanked towards the N-terminus by one or more residues. Other studies have also pointed to an important role for the peptide residue in the 6th position towards the C-terminus, relative to P1, for binding to various DR molecules.

Thus, peptides of the present invention are identified by any one of several HLA-specific amino acid motifs (*see, e.g.,* Tables I-III). If the presence of the motif corresponds to the ability to bind several allele-specific HLA antigens, it is referred to as a supermotif. The HLA molecules that bind to peptides that possess a particular amino acid supermotif are collectively referred to as an HLA "supertype."

The peptide motifs and supermotifs described below, and summarized in Tables I-III, provide guidance for the identification and use of peptide epitopes in accordance with the invention.

Examples of peptide epitopes bearing a respective supermotif or motif are included in Tables as designated in the description of each motif or supermotif below.

The Tables include a binding affinity ratio listing for some of the peptide epitopes. The ratio may be converted to IC₅₀ by using the following formula: IC₅₀ of the standard peptide/ratio = IC₅₀ of the test peptide (*i.e.,* the peptide epitope). The IC₅₀ values of standard peptides used to determine binding affinities for Class I peptides are shown in Table IV. The IC₅₀ values of standard peptides used to determine binding affinities for Class II peptides are shown in Table V. The peptides used as standards for the binding assays described herein are examples of standards; alternative standard peptides can also be used when performing such an analysis.

To obtain the peptide epitope sequences listed in each Table, protein sequence data from fourteen HCV isolates were evaluated for the presence of the designated supermotif or motif. The fourteen strains include HPCCGAA, HPCPLYPRE, HCV-H-CMR, HCV-J1, HPCGENANTI, HPCGENOM, HPCHUMR, HPCJCG, HPCJTA, HCV-J483, HCV-JK1, HCV-N, HPCPOLP, and HCV-J8. Peptide epitopes were additionally evaluated on the basis of their conservancy among these fourteen strains. A criterion for conservancy requires that the entire sequence of an HLA class I binding peptide be totally

conserved in 79% of the sequences available for a specific protein. Similarly, a criterion for conservancy requires that the entire 9-mer core region of an HLA class II binding peptide be totally conserved in 79% of the sequences available for a specific protein. The percent conservancy of the selected peptide epitopes is indicated on the Tables. The frequency, *i.e.* the number of strains of the fourteen strains in which the totally conserved peptide sequence was identified, is also shown. The "position" column in the Tables designates the amino acid position of the HCV polyprotein that corresponds to the first amino acid residue of the epitope. The "number of amino acids" indicates the number of residues in the epitope sequence.

HLA Class I Motifs Indicative of CTL Inducing Peptide Epitopes:

The primary anchor residues of the HLA class I peptide epitope supermotifs and motifs delineated below are summarized in Table I. The HLA class I motifs set out in Table I(a) are those most particularly relevant to the invention claimed here. Primary and secondary anchor positions are summarized in Table II. Allele-specific HLA molecules that comprise HLA class I supertype families are listed in Table VI.

IV.D.1. HLA-A1 supermotif

The HLA-A1 supermotif is characterized by the presence in peptide ligands of a small (T or S) or hydrophobic (L, I, V, or M) primary anchor residue in position 2, and an aromatic (Y, F, or W) primary anchor residue at the C-terminal position of the epitope. The corresponding family of HLA molecules that bind to the A1 supermotif (*i.e.*, the HLA-A1 supertype) is comprised of at least A*0101, A*2601, A*2602, A*2501, and A*3201 (*see, e.g.*, DiBrino, M. *et al.*, *J. Immunol.* 151:5930, 1993; DiBrino, M. *et al.*, *J. Immunol.* 152:620, 1994; Kondo, A. *et al.*, *Immunogenetics* 45:249, 1997). Other allele-specific HLA molecules predicted to be members of the A1 superfamily are shown in Table VI. Peptides binding to each of the individual HLA proteins can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the supermotif.

Representative peptide epitopes that comprise the A1 supermotif are set forth on the attached Table VII.

IV.D.2. HLA-A2 supermotif

Primary anchor specificities for allele-specific HLA-A2.1 molecules (Falk *et al.*, *Nature* 351:290-296, 1991; Hunt *et al.*, *Science* 255:1261-1263, 1992; Parker *et al.*, *J. Immunol.* 149:3580-3587, 1992) and cross-reactive binding within the HLA A2 family (Fruci *et al.*, *Human Immunol.* 38:187-192, 1993; Tanigaki *et al.*, *Human Immunol.* 39:155-162, 1994) have been described. The present inventors have defined additional primary anchor residues that determine cross-reactive binding to multiple allele-specific HLA A2 molecules (Ruppert *et al.*, *Cell* 74:929-937, 1993; Del Guercio *et al.*, *J. Immunol.* 154:685-693, 1995; Kast *et al.*, *J. Immunol.* 152:3904-3912, 1994). The HLA-A2 supermotif comprises peptide ligands with L, I, V, M, A, T, or Q as a primary anchor residue at position 2 and L, I, V, M, A, or T as a primary anchor residue at the C-terminal position of the epitope.

The corresponding family of HLA molecules (*i.e.*, the HLA-A2 supertype that binds these peptides) is comprised of at least: A*0201, A*0202, A*0203, A*0204, A*0205, A*0206, A*0207, A*0209, A*0214, A*6802, and A*6901. Other allele-specific HLA molecules predicted to be members of the A2 superfamily are shown in Table VI. As explained in detail below, binding to each of the individual allele-specific HLA molecules can be modulated by substitutions at the primary anchor and/or secondary anchor positions, preferably choosing respective residues specified for the supermotif.

Representative peptide epitopes that comprise an A2 supermotif are set forth on the attached Table VIII. The motifs comprising the primary anchor residues V, A, T, or Q at position 2 and L, I, V, A, or T at the C-terminal position are those most particularly relevant to the invention claimed herein.

IV.D.3. HLA-A3 supermotif

The HLA-A3 supermotif is characterized by the presence in peptide ligands of A, L, I, V, M, S, or, T as a primary anchor at position 2, and a positively charged residue, R or K, at the C-terminal position of the epitope (*e.g.*, in position 9 of 9-mers). Exemplary members of the corresponding family of HLA molecules (the HLA-A3 supertype) that bind the A3 supermotif include at least A*0301, A*1101, A*3101, A*3301, and A*6801. Other allele-specific HLA molecules predicted to be members of the A3 superfamily are shown in Table VI. As explained in detail below, peptide binding to each of the individual allele-specific HLA proteins can be modulated by substitutions of amino acids

at the primary and/or secondary anchor positions of the peptide, preferably choosing respective residues specified for the supermotif.

Representative peptide epitopes that comprise the A3 supermotif are set forth on the attached Table IX.

5

IV.D.4. HLA-A24 supermotif

The HLA-A24 supermotif is characterized by the presence in peptide ligands of an aromatic (F, W, or Y) or hydrophobic aliphatic (L, I, V, M, or T) residue as a primary anchor in position 2, and Y, F, W, L, I, or M as primary anchor at the C-terminal position of the epitope. The corresponding family of HLA molecules that bind to the A24 supermotif (*i.e.*, the A24 supertype) includes at least A*2402, A*3001, and A*2301. Other allele-specific HLA molecules predicted to be members of the A24 superfamily are shown in Table VI. Peptide binding to each of the allele-specific HLA molecules can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the supermotif.

15

Representative peptide epitopes that comprise the A24 supermotif are set forth on the attached Table X.

IV.D.5. HLA-B7 supermotif

The HLA-B7 supermotif is characterized by peptides bearing proline in position 2 as a primary anchor, and a hydrophobic or aliphatic amino acid (L, I, V, M, A, F, W, or Y) as the primary anchor at the C-terminal position of the epitope. The corresponding family of HLA molecules that bind the B7 supermotif (*i.e.*, the HLA-B7 supertype) is comprised of at least twenty six HLA-B proteins including: B*0702, B*0703, B*0704, B*0705, B*1508, B*3501, B*3502, B*3503, B*3504, B*3505, B*3506, B*3507, B*3508, B*5101, B*5102, B*5103, B*5104, B*5105, B*5301, B*5401, B*5501, B*5502, B*5601, B*5602, B*6701, and B*7801 (*see, e.g.*, Sidney, *et al.*, *J. Immunol.* 154:247, 1995; Barber, *et al.*, *Curr. Biol.* 5:179, 1995; Hill, *et al.*, *Nature* 360:434, 1992; Rammensee, *et al.*, *Immunogenetics* 41:178, 1995). Other allele-specific HLA molecules predicted to be members of the B7 superfamily are shown in Table VI. As explained in detail below, peptide binding to each of the individual allele-specific HLA proteins can be modulated by substitutions at the primary and/or secondary anchor positions of the peptide, preferably choosing respective residues specified for the supermotif.

20

25

30

Representative peptide epitopes that comprise the B7 supermotif are set forth on the attached Table XI.

IV.D.6. HLA-B27 supermotif

5 The HLA-B27 supermotif is characterized by the presence in peptide ligands of a positively charged (R, H, or K) residue as a primary anchor at position 2, and a hydrophobic (F, Y, L, W, M, I, A, or V) residue as a primary anchor at the C-terminal position of the epitope. Exemplary members of the corresponding family of HLA molecules that bind to the B27 supermotif (*i.e.*, the B27 supertype) include at least
10 B*1401, B*1402, B*1509, B*2702, B*2703, B*2704, B*2705, B*2706, B*3801, B*3901, B*3902, and B*7301. Other allele-specific HLA molecules predicted to be members of the B27 superfamily are shown in Table VI. Peptide binding to each of the allele specific HLA molecules can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the
15 supermotif.

Representative peptide epitopes that comprise the B27 supermotif are set forth on the attached Table XII.

IV.D.7. HLA-B44 supermotif

20 The HLA-B44 supermotif is characterized by the presence in peptide ligands of negatively charged (D or E) residues as a primary anchor in position 2, and hydrophobic residues (F, W, Y, L, I, M, V, or A) as a primary anchor at the C-terminal position of the epitope. Exemplary members of the corresponding family of HLA molecules that bind to the B44 supermotif (*i.e.*, the B44 supertype) include at least: B*1801, B*1802, B*3701,
25 B*4001, B*4002, B*4006, B*4402, B*4403, and B*4006. Peptide binding to each of the allele-specific HLA molecules can be modulated by substitutions at primary and/or secondary anchor positions; preferably choosing respective residues specified for the supermotif.

30 IV.D.8. HLA-B58 supermotif

The HLA-B58 supermotif is characterized by the presence in peptide ligands of a small aliphatic residue (A, S, or T) as a primary anchor residue at position 2, and an aromatic or hydrophobic residue (F, W, Y, L, I, V, M, or A) as a primary anchor residue at the C-terminal position of the epitope. Exemplary members of the corresponding

family of HLA molecules that bind to the B58 supermotif (*i.e.*, the B58 supertype) include at least: B*1516, B*1517, B*5701, B*5702, and B*5801. Other allele-specific HLA molecules predicted to be members of the B58 superfamily are shown in Table VI. Peptide binding to each of the allele-specific HLA molecules can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the supermotif.

Representative peptide epitopes that comprise the B58 supermotif are set forth on the attached Table XIII.

10 **IV.D.9. HLA-B62 supermotif**

The HLA-B62 supermotif is characterized by the presence in peptide ligands of the polar aliphatic residue Q or a hydrophobic aliphatic residue (L, V, M, I, or P) as a primary anchor in position 2, and a hydrophobic residue (F, W, Y, M, I, V, L, or A) as a primary anchor at the C-terminal position of the epitope. Exemplary members of the corresponding family of HLA molecules that bind to the B62 supermotif (*i.e.*, the B62 supertype) include at least: B*1501, B*1502, B*1513, and B5201. Other allele-specific HLA molecules predicted to be members of the B62 superfamily are shown in Table VI. Peptide binding to each of the allele-specific HLA molecules can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the supermotif.

Representative peptide epitopes that comprise the B62 supermotif are set forth on the attached Table XIV.

IV.D.10. HLA-A1 motif

25 The HLA-A1 motif is characterized by the presence in peptide ligands of T, S, or M as a primary anchor residue at position 2 and the presence of Y as a primary anchor residue at the C-terminal position of the epitope. An alternative allele-specific A1 motif is characterized by a primary anchor residue at position 3 rather than position 2. This motif is characterized by the presence of D, E, A, or S as a primary anchor residue in position 3, and a Y as a primary anchor residue at the C-terminal position of the epitope. Peptide binding to HLA A1 can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the motif.

Representative peptide epitopes that comprise either A1 motif are set forth on the attached Table XV. Those epitopes comprising T, S, or M at position 2 and Y at the C-terminal position are also included in the listing of HLA-A1 supermotif-bearing peptide epitopes listed in Table VII.

5

IV.D.11. HLA-A*0201 motif

An HLA-A2*0201 motif was first determined to be characterized by the presence in peptide ligands of L or M as a primary anchor residue in position 2, and L or V as a primary anchor residue at the C-terminal position of a 9-residue peptide (Falk *et al.*,
 10 *Nature* 351:290-296, 1991). The A*0201 motif was also determined to further comprise an I at position 2 and I or A at the C-terminal position of a nine amino acid peptide (Hunt *et al.*, *Science* 255:1261-1263, March 6, 1992; Parker *et al.*, *J. Immunol.* 149:3580-3587, 1992). Subsequently, the A*0201 allele-specific motif has been defined by the present inventors to additionally comprise V, A, T, or Q as a primary anchor residue at position 2,
 15 and M as a primary anchor residue at the C-terminal position of the epitope. Additionally, the A*0201 allele-specific motif has been found to comprise a T at the C-terminal position (Kast *et al.*, *J. Immunol.* 152:3904-3912, 1994). Thus, the HLA-A*0201 motif comprises peptide ligands with L, I, V, M, A, T, or Q as primary anchor residues at position 2 and L, I, V, M, A, or T as a primary anchor residue at the C-
 20 terminal position of the epitope. The preferred and tolerated residues that characterize the primary anchor positions of the HLA-A*0201 motif are identical to the residues describing the A2 supermotif. (For reviews of relevant data, *see, e.g.*, Del Guercio *et al.*, *J. Immunol.* 154:685-693, 1995; Ruppert *et al.*, *Cell* 74:929-937, 1993; Sidney *et al.*, *Immunol. Today* 17:261-266, 1996; Sette and Sidney, *Curr. Opin. in Immunol.* 10:478-
 25 482, 1998). Secondary anchor residues that characterize the A*0201 motif have additionally been defined as disclosed herein. These are disclosed in Table II. Peptide binding to HLA-A*0201 molecules can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the motif.

30 Representative peptide epitopes that comprise an A*0201 motif are set forth on the attached Table VIII. The A*0201 motifs comprising the primary anchor residues V, A, T, or Q at position 2 and L, I, V, A, or T at the C-terminal position are those most particularly relevant to the invention claimed herein.

IV.D.12. HLA-A3 motif

The HLA-A3 motif is characterized by the presence in peptide ligands of L, M, V, I, S, A, T, F, C, G, or D as a primary anchor residue at position 2, and the presence of K, Y, R, H, F, or A as a primary anchor residue at the C-terminal position of the epitope.

- 5 Peptide binding to HLA-A3 can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the motif.

Representative peptide epitopes that comprise the A3 motif are set forth on the attached Table XVI. Those peptide epitopes that also comprise the A3 supermotif are
10 also listed in Table IX.

IV.D.13. HLA-A11 motif

- The HLA-A11 motif is characterized by the presence in peptide ligands of V, I, M, L, I, S, A, G, N, C, D, or F as a primary anchor residue in position 2, and K, R, Y, or
15 H as a primary anchor residue at the C-terminal position of the epitope. Peptide binding to HLA-A11 can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the motif.

Representative peptide epitopes that comprise the A11 motif are set forth on the attached Table XVII; peptide epitopes comprising the A3 allele-specific motif are also
20 present in this Table because of the extensive overlap between the A3 and A11 motif primary anchor specificities. Further, those peptide epitopes that comprise the A3 supermotif are also listed in Table IX.

IV.D.14. HLA-A24 motif

- 25 The HLA-A24 motif is characterized by the presence in peptide ligands of Y, F, W, or M as a primary anchor residue in position 2, and F, L, I, or W as a primary anchor residue at the C-terminal position of the epitope. Peptide binding to HLA-A24 molecules can be modulated by substitutions at primary and/or secondary anchor positions; preferably choosing respective residues specified for the motif.

30 Representative peptide epitopes that comprise the A24 motif are set forth on the attached Table XVIII. These epitopes are also listed in Table X, which sets forth HLA-A24-supermotif-bearing peptide epitopes.

Motifs Indicative of Class II HTL Inducing Peptide Epitopes

The primary and secondary anchor residues of the HLA class II peptide epitope supermotifs and motifs delineated below are summarized in Table III.

5 IV.D.15. HLA DR-1-4-7 supermotif

Motifs have also been identified for peptides that bind to three common HLA class II allele-specific HLA molecules: HLA DRB1*0401, DRB1*0101, and DRB1*0701. Collectively, the common residues from these motifs delineate the HLA DR-1-4-7 supermotif. Peptides that bind to these DR molecules carry a supermotif
 10 characterized by a large aromatic or hydrophobic residue (Y, F, W, L, I, V, or M) as a primary anchor residue in position 1, and a small, non-charged residue (S, T, C, A, P, V, I, L, or M) as a primary anchor residue in position 6 of a 9-mer core region. Allele-specific secondary effects and secondary anchors for each of these HLA types have also been identified. These are set forth in Table III. Peptide binding to HLA- DRB1*0401,
 15 DRB1*0101, and/or DRB1*0701 can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the supermotif.

Conserved peptide epitopes *i.e.*, conserved in $\geq 79\%$ ($\geq 11/14$) of the HCV strains used for the present analysis, may be described as corresponding to epitopes containing a
 20 nine residue core comprising the DR-1-4-7 supermotif, and in which the 9 residue core is conserved in $\geq 79\%$ (wherein position 1 of the motif is at position 1 of the nine residue core). Conserved 9-mer core regions are set forth in Table XIXa. Respective exemplary peptide epitopes of 15 amino acid residues in length, each of which comprise a conserved nine residue core, are also shown in section "a" of the Table. Cross-reactive binding data
 25 for exemplary 15-residue supermotif-bearing peptides are shown in Table XIXb.

IV.D.16. HLA DR3 motifs

Two alternative motifs (*i.e.*, submotifs) characterize peptide epitopes that bind to HLA-DR3 molecules. In the first motif (submotif DR3A) a large, hydrophobic residue
 30 (L, I, V, M, F, or Y) is present in anchor position 1 of a 9-mer core, and D is present as an anchor at position 4, towards the carboxyl terminus of the epitope. As in other class II motifs, core position 1 may or may not occupy the peptide N-terminal position.

The alternative DR3 submotif provides for lack of the large, hydrophobic residue at anchor position 1, and/or lack of the negatively charged or amide-like anchor residue at position 4, by the presence of a positive charge at position 6 towards the carboxyl terminus of the epitope. Thus, for the alternative allele-specific DR3 motif (submotif DR3B): L, I, V, M, F, Y, A, or Y is present at anchor position 1; D, N, Q, E, S, or T is present at anchor position 4; and K, R, or H is present at anchor position 6. Peptide binding to HLA-DR3 can be modulated by substitutions at primary and/or secondary anchor positions, preferably choosing respective residues specified for the motif.

Conserved 9-mer core regions (*i.e.*, those sequences that are conserved in at least 79% of the 14 HCV strains used for the analysis) corresponding to a nine residue sequence comprising the DR3A submotif (wherein position 1 of the motif is at position 1 of the nine residue core) are set forth in Table XXa. Respective exemplary peptide epitopes of 15 amino acid residues in length, each of which comprise a conserved nine residue core, are also shown in Table XXa. Table XXb shows binding data of exemplary DR3 submotif A-bearing peptides.

Conserved 9-mer core regions (*i.e.*, those that are at least 79% conserved in the 14 HCV strains used for the analysis) comprising the DR3B submotif and respective exemplary 15-mer peptides comprising the DR3 submotif-B epitope are set forth in Table XXc. Table XXd shows binding data of exemplary DR3 submotif B-bearing peptides.

Each of the HLA class I or class II peptide epitopes set out in the Tables herein are deemed singly to be an inventive aspect of this application. Further, it is also an inventive aspect of this application that each peptide epitope may be used in combination with any other peptide epitope.

IV.E. Enhancing Population Coverage of the Vaccine

Vaccines that have broad population coverage are preferred because they are more commercially viable and generally applicable to the most people. Broad population coverage can be obtained using the peptides of the invention (and nucleic acid compositions that encode such peptides) through selecting peptide epitopes that bind to HLA alleles which, when considered in total, are present in most of the population. Table XXI lists the overall frequencies of the HLA class I supertypes in various ethnicities (Table XXIa) and the combined population coverage achieved by the A2-, A3-, and B7-supertypes (Table XXIb). The A2-, A3-, and B7 supertypes are each present on the average of over 40% in each of these five major ethnic groups. Coverage in excess of

80% is achieved with a combination of these supermotifs. These results suggest that effective and non-ethnically biased population coverage is achieved upon use of a limited number of cross-reactive peptides. Although the population coverage reached with these three main peptide specificities is high, coverage can be expanded to reach 95%

5 population coverage and above, and more easily achieve truly multispecific responses upon use of additional supermotif or allele-specific motif bearing peptides.

The B44-, A1-, and A24-supertypes are present, on average, in a range from 25% to 40% in these major ethnic populations (Table XXIa). While less prevalent overall, the B27-, B58-, and B62 supertypes are each present with a frequency >25% in at least one
10 major ethnic group (Table XXIa). Table XXIb summarizes the estimated prevalence of combinations of HLA supertypes that have been identified in five major ethnic groups. The incremental coverage obtained by the inclusion of A1-, A24-, and B44-supertypes to the A2, A3, and B7 coverage, or all of the supertypes described herein, is shown.

The data presented herein, together with the previous definition of the A2-, A3-,
15 and B7-supertypes, indicates that all antigens, with the possible exception of A29, B8, and B46, can be classified into a total of nine HLA supertypes. By including epitopes from the six most frequent supertypes, an average population coverage of 99% is obtained for five major ethnic groups..

20 **IV.F. Immune Response-Stimulating Peptide Analogs**

In general, CTL and HTL responses are not directed against all possible epitopes. Rather, they are restricted to a few "immunodominant" determinants (Zinkernagel, *et al.*, *Adv. Immunol.* 27:5159, 1979; Bennink, *et al.*, *J. Exp. Med.* 168:1935-1939, 1988; Rawle, *et al.*, *J. Immunol.* 146:3977-3984, 1991). It has been recognized that immunodominance
25 (Benacerraf, *et al.*, *Science* 175:273-279, 1972) could be explained by either the ability of a given epitope to selectively bind a particular HLA protein (determinant selection theory) (Vitiello, *et al.*, *J. Immunol.* 131:1635, 1983); Rosenthal, *et al.*, *Nature* 267:156-158, 1977), or to be selectively recognized by the existing TCR (T cell receptor) specificities (repertoire theory) (Klein, J., *IMMUNOLOGY, THE SCIENCE OF SELF/NONSELF*
30 *DISCRIMINATION*, John Wiley & Sons, New York, pp. 270-310, 1982). It has been demonstrated that additional factors, mostly linked to processing events, can also play a key role in dictating, beyond strict immunogenicity, which of the many potential determinants will be presented as immunodominant (Sercarz, *et al.*, *Annu. Rev. Immunol.* 11:729-766, 1993).

The concept of dominance and subdominance is relevant to immunotherapy of both infectious diseases and cancer. For example, in the course of chronic viral disease, recruitment of subdominant epitopes can be important for successful clearance of the infection, especially if dominant CTL or HTL specificities have been inactivated by functional tolerance, suppression, mutation of viruses and other mechanisms (Franco, *et al.*, *Curr. Opin. Immunol.* 7:524-531, 1995). In the case of cancer and tumor antigens, CTLs recognizing at least some of the highest binding affinity peptides might be functionally inactivated. Lower binding affinity peptides are preferentially recognized at these times, and may therefore be preferred in therapeutic or prophylactic anti-cancer vaccines.

In particular, it has been noted that a significant number of epitopes derived from known non-viral tumor associated antigens (TAA) bind HLA class I with intermediate affinity (IC_{50} in the 50-500 nM range). For example, it has been found that 8 of 15 known TAA peptides recognized by tumor infiltrating lymphocytes (TIL) or CTL bound in the 50-500 nM range. (These data are in contrast with estimates that 90% of known viral antigens were bound by HLA class I molecules with IC_{50} of 50 nM or less, while only approximately 10% bound in the 50-500 nM range (Sette, *et al.*, *J. Immunol.*, 153:558-5592, 1994). In the cancer setting this phenomenon is probably due to elimination or functional inhibition of the CTL recognizing several of the highest binding peptides, presumably because of T cell tolerization events.

Without intending to be bound by theory, it is believed that because T cells to dominant epitopes may have been clonally deleted, selecting subdominant epitopes may allow existing T cells to be recruited, which will then lead to a therapeutic or prophylactic response. However, the binding of HLA molecules to subdominant epitopes is often less vigorous than to dominant ones. Accordingly, there is a need to be able to modulate the binding affinity of particular immunogenic epitopes for one or more HLA molecules, and thereby to modulate the immune response elicited by the peptide, for example to prepare analog peptides which elicit a more vigorous response. This ability would greatly enhance the usefulness of peptide-based vaccines and therapeutic agents.

Although peptides with suitable cross-reactivity among all alleles of a superfamily are identified by the screening procedures described above, cross-reactivity is not always as complete as possible, and in certain cases procedures to increase cross-reactivity of peptides can be useful; moreover, such procedures can also be used to modify other properties of the peptides such as binding affinity or peptide stability. Having established

the general rules that govern cross-reactivity of peptides for HLA alleles within a given motif or supermotif, modification (*i.e.*, analoging) of the structure of peptides of particular interest in order to achieve broader (or otherwise modified) HLA binding capacity can be performed. More specifically, peptides which exhibit the broadest cross-reactivity patterns, can be produced in accordance with the teachings herein. The present concepts related to analog generation are set forth in greater detail in co-pending U.S.S.N. 09/226,775 filed 1/6/99.

In brief, the strategy employed utilizes the motifs or supermotifs which correlate with binding to certain HLA molecules. The motifs or supermotifs are defined by having primary anchors, and in many cases secondary anchors. Analog peptides can be created by substituting amino acid residues at primary anchor, secondary anchor, or at primary and secondary anchor positions. Generally, analogs are made for peptides that already bear a motif or supermotif. Preferred secondary anchor residues of supermotifs and motifs that have been defined for HLA class I and class II binding peptides are shown in Tables II and III, respectively.

For a number of the motifs or supermotifs in accordance with the invention, residues are defined which are deleterious to binding to allele-specific HLA molecules or members of HLA supertypes that bind the respective motif or supermotif (Tables II and III). Accordingly, removal of such residues that are detrimental to binding can be performed in accordance with the present invention. For example, in the case of the A3 supertype, when all peptides that have such deleterious residues are removed from the population of analyzed peptides, the incidence of cross-reactivity increases from 22% to 37% (*see, e.g.*, Sidney, J. *et al.*, *Hu. Immunol.* 45:79, 1996). Thus, one strategy to improve the cross-reactivity of peptides within a given supermotif is simply to delete one or more of the deleterious residues present within a peptide and substitute a small "neutral" residue such as Ala (that may not influence T cell recognition of the peptide). An enhanced likelihood of cross-reactivity is expected if, together with elimination of detrimental residues within a peptide, "preferred" residues associated with high affinity binding to an allele-specific HLA molecule or to multiple HLA molecules within a superfamily are inserted.

To ensure that an analog peptide, when used as a vaccine, actually elicits a CTL response to the native epitope *in vivo* (or, in the case of class II epitopes, elicits helper T cells that cross-react with the wild type peptides), the analog peptide may be used to immunize T cells *in vitro* from individuals of the appropriate HLA allele. Thereafter, the

immunized cells' capacity to induce lysis of wild type peptide sensitized target cells is evaluated. It will be desirable to use as antigen presenting cells, cells that have been either infected, or transfected with the appropriate genes, or, in the case of class II epitopes only, cells that have been pulsed with whole protein antigens, to establish
 5 whether endogenously produced antigen is also recognized by the relevant T cells.

Another embodiment of the invention is to create analogs of weak binding peptides, to thereby ensure adequate numbers of cross-reactive cellular binders. Class I binding peptides exhibiting binding affinities of 500-5000 nM, and carrying an acceptable but suboptimal primary anchor residue at one or both positions can be "fixed" by
 10 substituting preferred anchor residues in accordance with the respective supertype. The analog peptides can then be tested for crossbinding activity.

Another embodiment for generating effective peptide analogs involves the substitution of residues that have an adverse impact on peptide stability or solubility in, *e.g.*, a liquid environment. This substitution may occur at any position of the peptide
 15 epitope. For example, a cysteine (C) can be substituted out in favor of α -amino butyric acid. Due to its chemical nature, cysteine has the propensity to form disulfide bridges and sufficiently alter the peptide structurally so as to reduce binding capacity. Substituting α -amino butyric acid for C not only alleviates this problem, but actually improves binding and crossbinding capability in certain instances (*see, e.g.*, the review by Sette *et al.*, In:
 20 Persistent Viral Infections, Eds. R. Ahmed and I. Chen, John Wiley & Sons, England, 1999). Substitution of cysteine with α -amino butyric acid may occur at any residue of a peptide epitope, *i.e.* at either anchor or non-anchor positions.

Representative analog peptides are set forth in Table XXII. The Table indicates the length and sequence of the analog peptide as well as the motif or supermotif, if
 25 appropriate. The information in the "Fixed Nomenclature" column indicates the residues substituted at the indicated position numbers for the respective analog.

IV.G. Computer Screening of Protein Sequences from Disease-Related Antigens for Supermotif- or Motif-Bearing Peptides

30 In order to identify supermotif- or motif-bearing epitopes in a target antigen, a native protein sequence, *e.g.*, a tumor-associated antigen, or sequences from an infectious organism, or a donor tissue for transplantation, is screened using a means for computing, such as an intellectual calculation or a computer, to determine the presence of a supermotif or motif within the sequence. The information obtained from the analysis of

native peptide can be used directly to evaluate the status of the native peptide or may be utilized subsequently to generate the peptide epitope.

Computer programs that allow the rapid screening of protein sequences for the occurrence of the subject supermotifs or motifs are encompassed by the present invention; as are programs that permit the generation of analog peptides. These programs are implemented to analyze any identified amino acid sequence or operate on an unknown sequence and simultaneously determine the sequence and identify motif-bearing epitopes thereof; analogs can be simultaneously determined as well. Generally, the identified sequences will be from a pathogenic organism or a tumor-associated peptide. For example, the target molecules considered herein include, without limitation, the core, S, E1, NS1/E2, NS2, NS3, NS4, and NS5 regions of HCV.

In cases where the sequence of multiple variants of the same target protein are available, peptides may also be selected on the basis of their conservancy. A presently preferred criterion for conservancy defines that the entire sequence of an HLA class I binding peptide or the entire 9-mer core of a class II binding peptide, be totally (*i.e.*, 100%) conserved in at least 79% of the sequences evaluated for a specific protein. This definition of conservancy has been employed herein; although, as appreciated by those in the art, lower or higher degrees of conservancy can be employed as appropriate for a given antigenic target.

It is important that the selection criteria utilized for prediction of peptide binding are as accurate as possible, to correlate most efficiently with actual binding. Prediction of peptides that bind, for example, to HLA-A*0201, on the basis of the presence of the appropriate primary anchors, is positive at about a 30% rate (see, *e.g.*, Ruppert, J. *et al. Cell* 74:929, 1993). However, by extensively analyzing peptide-HLA binding data disclosed herein, data in related patent applications, and data in the art, the present inventors have developed a number of allele-specific polynomial algorithms that dramatically increase the predictive value over identification on the basis of the presence of primary anchor residues alone. These algorithms take into account not only the presence or absence of primary anchors, but also consider the positive or deleterious presence of secondary anchor residues (to account for the impact of different amino acids at different positions). The algorithms are essentially based on the premise that the overall affinity (or ΔG) of peptide-HLA interactions can be approximated as a linear polynomial function of the type:

$$\Delta G = a_{1i} \times a_{2i} \times a_{3i} \dots \times a_{ni}$$

where a_{ji} is a coefficient that represents the effect of the presence of a given amino acid (j) at a given position (i) along the sequence of a peptide of n amino acids. An important assumption of this method is that the effects at each position are essentially independent of each other. This assumption is justified by studies that demonstrated that peptides are bound to HLA molecules and recognized by T cells in essentially an extended conformation. Derivation of specific algorithm coefficients has been described, for example, in Gulukota, K. *et al.*, *J. Mol. Biol.* 267:1258, 1997.

Additional methods to identify preferred peptide sequences, which also make use of specific motifs, include the use of neural networks and molecular modeling programs (see, e.g., Milik *et al.*, *Nature Biotechnology* 16:753, 1998; Altuvia *et al.*, *Hum. Immunol.* 58:1, 1997; Altuvia *et al.*, *J. Mol. Biol.* 249:244, 1995; Buus, S. *Curr. Opin. Immunol.* 11:209-213, 1999; Brusic, V. *et al.*, *Bioinformatics* 14:121-130, 1998; Parker *et al.*, *J. Immunol.* 152:163, 1993; Meister *et al.*, *Vaccine* 13:581, 1995; Hammer *et al.*, *J. Exp. Med.* 180:2353, 1994; Sturniolo *et al.*, *Nature Biotechnol.* 17:555 1999).

For example, it has been shown that in sets of A*0201 motif-bearing peptides containing at least one preferred secondary anchor residue while avoiding the presence of any deleterious secondary anchor residues, 69% of the peptides will bind A*0201 with an IC_{50} less than 500 nM (Ruppert, J. *et al.* *Cell* 74:929, 1993). These algorithms are also flexible in that cut-off scores may be adjusted to select sets of peptides with greater or lower predicted binding properties, as desired.

In utilizing computer screening to identify peptide epitopes, a protein sequence or translated sequence may be analyzed using software developed to search for motifs, for example the "FINDPATTERNS" program (Devereux, *et al.* *Nucl. Acids Res.* 12:387-395, 1984) or MotifSearch 1.4 software program (D. Brown, San Diego, CA) to identify potential peptide sequences containing appropriate HLA binding motifs. The identified peptides can be scored using customized polynomial algorithms to predict their capacity to bind specific HLA class I or class II alleles. As appreciated by one of ordinary skill in the art, a large array of computer programming software and hardware options are available in the relevant art which can be employed to implement the motifs of the invention in order to evaluate (e.g., without limitation, to identify epitopes, identify epitope concentration per peptide length, or to generate analogs) known or unknown peptide sequences.

In accordance with the procedures described above, HCV peptide epitopes and analogs thereof that are able to bind HLA supertype groups or allele-specific HLA molecules have been identified (Tables VII-XX; Table XXII).

5 **IV.H. Preparation of Peptide Epitopes**

Peptides in accordance with the invention can be prepared synthetically, by recombinant DNA technology or chemical synthesis, or from natural sources such as native tumors or pathogenic organisms. Peptide epitopes may be synthesized individually or as polyepitopic peptides. Although the peptide will preferably be substantially free of
10 other naturally occurring host cell proteins and fragments thereof, in some embodiments the peptides may be synthetically conjugated to native fragments or particles.

The peptides in accordance with the invention can be a variety of lengths, and either in their neutral (uncharged) forms or in forms which are salts. The peptides in accordance with the invention are either free of modifications such as glycosylation, side
15 chain oxidation, or phosphorylation; or they contain these modifications, subject to the condition that modifications do not destroy the biological activity of the peptides as described herein.

Desirably, the peptide epitope will be as small as possible while still maintaining substantially all of the immunologic activity of the native protein. When possible, it may
20 be desirable to optimize HLA class I binding peptide epitopes of the invention to a length of about 8 to about 13 amino acid residues, preferably 9 to 10. HLA class II binding peptide epitopes may be optimized to a length of about 6 to about 30 amino acids in length, preferably to between about 13 and about 20 residues. Preferably, the peptide epitopes are commensurate in size with endogenously processed pathogen-derived
25 peptides or tumor cell peptides that are bound to the relevant HLA molecules.

The identification and preparation of peptides of other lengths can also be carried out using the techniques described herein. Moreover, it is preferred to identify native peptide regions that contain a high concentration of class I and/or class II epitopes. Such a sequence is generally selected on the basis that it contains the greatest number of
30 epitopes per amino acid length. It is to be appreciated that epitopes can be present in a frame-shifted manner, *e.g.* a 10 amino acid long peptide could contain two 9 amino acid long epitopes and one 10 amino acid long epitope; upon intracellular processing, each epitope can be exposed and bound by an HLA molecule upon administration of such a

peptide. This larger, preferably multi-epitopic, peptide can be generated synthetically, recombinantly, or via cleavage from the native source.

The peptides of the invention can be prepared in a wide variety of ways. For the preferred relatively short size, the peptides can be synthesized in solution or on a solid support in accordance with conventional techniques. Various automatic synthesizers are commercially available and can be used in accordance with known protocols. (See, for example, Stewart & Young, SOLID PHASE PEPTIDE SYNTHESIS, 2D. ED., Pierce Chemical Co., 1984). Further, individual peptide epitopes can be joined using chemical ligation to produce larger peptides that are still within the bounds of the invention.

Alternatively, recombinant DNA technology can be employed wherein a nucleotide sequence which encodes an immunogenic peptide of interest is inserted into an expression vector, transformed or transfected into an appropriate host cell and cultivated under conditions suitable for expression. These procedures are generally known in the art, as described generally in Sambrook *et al.*, MOLECULAR CLONING, A LABORATORY MANUAL, Cold Spring Harbor Press, Cold Spring Harbor, New York (1989). Thus, recombinant polypeptides which comprise one or more peptide sequences of the invention can be used to present the appropriate T cell epitope.

The nucleotide coding sequence for peptide epitopes of the preferred lengths contemplated herein can be synthesized by chemical techniques, for example, the phosphotriester method of Matteucci, *et al.*, *J. Am. Chem. Soc.* 103:3185 (1981). Peptide analogs can be made simply by substituting the appropriate and desired nucleic acid base(s) for those that encode the native peptide sequence; exemplary nucleic acid substitutions are those that encode an amino acid defined by the motifs/super-motifs herein. The coding sequence can then be provided with appropriate linkers and ligated into expression vectors commonly available in the art, and the vectors used to transform suitable hosts to produce the desired fusion protein. A number of such vectors and suitable host systems are now available. For expression of the fusion proteins, the coding sequence will be provided with operably linked start and stop codons, promoter and terminator regions and usually a replication system to provide an expression vector for expression in the desired cellular host. For example, promoter sequences compatible with bacterial hosts are provided in plasmids containing convenient restriction sites for insertion of the desired coding sequence. The resulting expression vectors are transformed into suitable bacterial hosts. Of course, yeast, insect or mammalian cell hosts may also be used, employing suitable vectors and control sequences.

IV.I. Assays to Detect T-Cell Responses

Once HLA binding peptides are identified, they can be tested for the ability to elicit a T-cell response. The preparation and evaluation of motif-bearing peptides are described in PCT publications WO 94/20127 and WO 94/03205. Briefly, peptides comprising epitopes from a particular antigen are synthesized and tested for their ability to bind to the appropriate HLA proteins. These assays may involve evaluating the binding of a peptide of the invention to purified HLA class I molecules in relation to the binding of a radioiodinated reference peptide. Alternatively, cells expressing empty class I molecules (*i.e.* lacking peptide therein) may be evaluated for peptide binding by immunofluorescent staining and flow microfluorimetry. Other assays that may be used to evaluate peptide binding include peptide-dependent class I assembly assays and/or the inhibition of CTL recognition by peptide competition. Those peptides that bind to the class I molecule, typically with an affinity of 500 nM or less, are further evaluated for their ability to serve as targets for CTLs derived from infected or immunized individuals, as well as for their capacity to induce primary *in vitro* or *in vivo* CTL responses that can give rise to CTL populations capable of reacting with selected target cells associated with a disease. Corresponding assays are used for evaluation of HLA class II binding peptides. HLA class II motif-bearing peptides that are shown to bind, typically at an affinity of 1000 nM or less, are further evaluated for the ability to stimulate HTL responses.

Conventional assays utilized to detect T cell responses include proliferation assays, lymphokine secretion assays, direct cytotoxicity assays, and limiting dilution assays. For example, antigen-presenting cells that have been incubated with a peptide can be assayed for the ability to induce CTL responses in responder cell populations. Antigen-presenting cells can be normal cells such as peripheral blood mononuclear cells or dendritic cells. Alternatively, mutant non-human mammalian cell lines that are deficient in their ability to load class I molecules with internally processed peptides and that have been transfected with the appropriate human class I gene, may be used to test for the capacity of the peptide to induce *in vitro* primary CTL responses.

Peripheral blood mononuclear cells (PBMCs) may be used as the responder cell source of CTL precursors. The appropriate antigen-presenting cells are incubated with peptide, after which the peptide-loaded antigen-presenting cells are then incubated with the responder cell population under optimized culture conditions. Positive CTL

activation can be determined by assaying the culture for the presence of CTLs that kill radio-labeled target cells, both specific peptide-pulsed targets as well as target cells expressing endogenously processed forms of the antigen from which the peptide sequence was derived.

- 5 More recently, a method has been devised which allows direct quantification of antigen-specific T cells by staining with Fluorescein-labelled HLA tetrameric complexes (Altman, J. D. *et al.*, *Proc. Natl. Acad. Sci. USA* 90:10330, 1993; Altman, J. D. *et al.*, *Science* 274:94, 1996). Other relatively recent technical developments include staining for intracellular lymphokines, and interferon release assays or ELISPOT assays.
- 10 Tetramer staining, intracellular lymphokine staining and ELISPOT assays all appear to be at least 10-fold more sensitive than more conventional assays (Lalvani, A. *et al.*, *J. Exp. Med.* 186:859, 1997; Dunbar, P. R. *et al.*, *Curr. Biol.* 8:413, 1998; Murali-Krishna, K. *et al.*, *Immunity* 8:177, 1998).

- HTL activation may also be assessed using such techniques known to those in the art such as T cell proliferation and secretion of lymphokines, *e.g.* IL-2 (*see, e.g.* Alexander *et al.*, *Immunity* 1:751-761, 1994).
- 15

- Alternatively, immunization of HLA transgenic mice can be used to determine immunogenicity of peptide epitopes. Several transgenic mouse models including mice with human A2.1, A11 (which can additionally be used to analyze HLA-A3 epitopes), and B7 alleles have been characterized and others (*e.g.*, transgenic mice for HLA-A1 and A24) are being developed. HLA-DR1 and HLA-DR3 mouse models have also been developed. Additional transgenic mouse models with other HLA alleles may be generated as necessary. Mice may be immunized with peptides emulsified in Incomplete Freund's Adjuvant and the resulting T cells tested for their capacity to recognize peptide-pulsed target cells and target cells transfected with appropriate genes. CTL responses may be analyzed using cytotoxicity assays described above. Similarly, HTL responses may be analyzed using such assays as T cell proliferation or secretion of lymphokines.
- 20
- 25

Exemplary immunogenic peptide epitopes are set out in Table XXIII.

30 **IV.J. Use of Peptide Epitopes as Diagnostic Agents and for Evaluating Immune Responses**

HLA class I and class II binding peptides as described herein can be used, in one embodiment of the invention, as reagents to evaluate an immune response. The immune response to be evaluated may be induced by using as an immunogen any agent that may

result in the production of antigen-specific CTLs or HTLs that recognize and bind to the peptide epitope(s) to be employed as the reagent. The peptide reagent need not be used as the immunogen. Assay systems that may be used for such an analysis include relatively recent technical developments such as tetramers, staining for intracellular lymphokines and interferon release assays, or ELISPOT assays.

For example, a peptide of the invention may be used in a tetramer staining assay to assess peripheral blood mononuclear cells for the presence of antigen-specific CTLs following exposure to a pathogen or immunogen. The HLA-tetrameric complex is used to directly visualize antigen-specific CTLs (*see, e.g., Ogg et al., Science* 279:2103-2106, 1998; and Altman *et al., Science* 174:94-96, 1996) and determine the frequency of the antigen-specific CTL population in a sample of peripheral blood mononuclear cells. A tetramer reagent using a peptide of the invention may be generated as follows: A peptide that binds to an HLA molecule is refolded in the presence of the corresponding HLA heavy chain and β_2 -microglobulin to generate a trimolecular complex. The complex is biotinylated at the carboxyl terminal end of the heavy chain at a site that was previously engineered into the protein. Tetramer formation is then induced by the addition of streptavidin. By means of fluorescently labeled streptavidin, the tetramer can be used to stain antigen-specific cells. The cells may then be identified, for example, by flow cytometry. Such an analysis may be used for diagnostic or prognostic purposes.

Peptides of the invention may also be used as reagents to evaluate immune recall responses. (*see, e.g., Bertonni et al., J. Clin. Invest.* 100:503-513, 1997 and Penna *et al., J. Exp. Med.* 174:1565-1570, 1991.) For example, patient PBMC samples from individuals with HCV infection may be analyzed for the presence of antigen-specific CTLs or HTLs using specific peptides. A blood sample containing mononuclear cells may be evaluated by cultivating the PBMCs and stimulating the cells with a peptide of the invention. After an appropriate cultivation period, the expanded cell population may be analyzed, for example, for cytotoxic activity (CTL) or for HTL activity.

The peptides may also be used as reagents to evaluate the efficacy of a vaccine. PBMCs obtained from a patient vaccinated with an immunogen may be analyzed using, for example, either of the methods described above. The patient is HLA typed, and peptide epitope reagents that recognize the allele-specific molecules present in that patient are selected for the analysis. The immunogenicity of the vaccine is indicated by the presence of HCV epitope-specific CTLs and/or HTLs in the PBMC sample.

The peptides of the invention may also be used to make antibodies, using techniques well known in the art (see, e.g. *CURRENT PROTOCOLS IN IMMUNOLOGY*, Wiley/Greene, NY; and *Antibodies A Laboratory Manual* Harlow, Harlow and Lane, Cold Spring Harbor Laboratory Press, 1989), which may be useful as reagents to
 5 diagnose HCV infection. Such antibodies include those that recognize a peptide in the context of an HLA molecule, i.e., antibodies that bind to a peptide-MHC complex.

IV.K. Vaccine Compositions

Vaccines that contain an immunogenically effective amount of one or more
 10 peptides as described herein are a further embodiment of the invention. Once appropriately immunogenic epitopes have been defined, they can be sorted and delivered by various means, herein referred to as "vaccine" compositions. Such vaccine compositions can include, for example, lipopeptides (e.g., Vitiello, A. *et al.*, *J. Clin. Invest.* 95:341, 1995), peptide compositions encapsulated in poly(DL-lactide-co-
 15 glycolide) ("PLG") microspheres (see, e.g., Eldridge, *et al.*, *Molec. Immunol.* 28:287-294, 1991; Alonso *et al.*, *Vaccine* 12:299-306, 1994; Jones *et al.*, *Vaccine* 13:675-681, 1995), peptide compositions contained in immune stimulating complexes (ISCOMS) (see, e.g., Takahashi *et al.*, *Nature* 344:873-875, 1990; Hu *et al.*, *Clin Exp Immunol.* 113:235-243, 1998), multiple antigen peptide systems (MAPs) (see e.g., Tam, J. P., *Proc. Natl. Acad.*
 20 *Sci. U.S.A.* 85:5409-5413, 1988; Tam, J.P., *J. Immunol. Methods* 196:17-32, 1996), viral delivery vectors (Perkus, M. E. *et al.*, In: *Concepts in vaccine development*, Kaufmann, S. H. E., ed., p. 379, 1996; Chakrabarti, S. *et al.*, *Nature* 320:535, 1986; Hu, S. L. *et al.*, *Nature* 320:537, 1986; Kieny, M.-P. *et al.*, *AIDS Bio/Technology* 4:790, 1986; Top, F. H. *et al.*, *J. Infect. Dis.* 124:148, 1971; Chanda, P. K. *et al.*, *Virology* 175:535, 1990),
 25 particles of viral or synthetic origin (e.g., Kofler, N. *et al.*, *J. Immunol. Methods* 192:25, 1996; Eldridge, J. H. *et al.*, *Sem. Hematol.* 30:16, 1993; Falo, L. D., Jr. *et al.*, *Nature Med.* 7:649, 1995), adjuvants (Warren, H. S., Vogel, F. R., and Chedid, L. A. *Annu. Rev. Immunol.* 4:369, 1986; Gupta, R. K. *et al.*, *Vaccine* 11:293, 1993), liposomes (Reddy, R. *et al.*, *J. Immunol.* 148:1585, 1992; Rock, K. L., *Immunol. Today* 17:131, 1996), or,
 30 naked or particle absorbed cDNA (Ulmer, J. B. *et al.*, *Science* 259:1745, 1993; Robinson, H. L., Hunt, L. A., and Webster, R. G., *Vaccine* 11:957, 1993; Shiver, J. W. *et al.*, In: *Concepts in vaccine development*, Kaufmann, S. H. E., ed., p. 423, 1996; Cease, K. B., and Berzofsky, J. A., *Annu. Rev. Immunol.* 12:923, 1994 and Eldridge, J. H. *et al.*, *Sem. Hematol.* 30:16, 1993). Toxin-targeted delivery technologies, also known as receptor

mediated targeting, such as those of Avant Immunotherapeutics, Inc. (Needham, Massachusetts) may also be used.

Furthermore, vaccines in accordance with the invention encompass compositions of one or more of the claimed peptide(s). The peptide(s) can be individually linked to its own carrier; alternatively, the peptide(s) can exist as a homopolymer or heteropolymer of active peptide units. Such a polymer has the advantage of increased immunological reaction and, where different peptide epitopes are used to make up the polymer, the additional ability to induce antibodies and/or CTLs that react with different antigenic determinants of the pathogenic organism or tumor-related peptide targeted for an immune response. The composition may be a naturally occurring region of an antigen or may be prepared, *e.g.*, recombinantly or by chemical synthesis.

Furthermore, useful carriers that can be used with vaccines of the invention are well known in the art, and include, *e.g.*, thyroglobulin, albumins such as human serum albumin, tetanus toxoid, polyamino acids such as poly L-lysine, poly L-glutamic acid, influenza, hepatitis B virus core protein, and the like. The vaccines can contain a physiologically tolerable (*i.e.*, acceptable) diluent such as water, or saline, preferably phosphate buffered saline. The vaccines also typically include an adjuvant. Adjuvants such as incomplete Freund's adjuvant, aluminum phosphate, aluminum hydroxide, or alum are examples of materials well known in the art. Additionally, as disclosed herein, CTL responses can be primed by conjugating peptides of the invention to lipids, such as tripalmitoyl-S-glycerylcysteinylserine (P₃CSS).

As disclosed in greater detail herein, upon immunization with a peptide composition in accordance with the invention, via injection, aerosol, oral, transdermal, transmucosal, intrapleural, intrathecal, or other suitable routes, the immune system of the host responds to the vaccine by producing large amounts of CTLs specific for the desired antigen. Consequently, the host becomes at least partially immune to later infection, or at least partially resistant to developing an ongoing chronic infection, or derives at least some therapeutic benefit when the antigen was tumor-associated.

In some instances it may be desirable to combine the class I peptide vaccines of the invention with vaccines which induce or facilitate neutralizing antibody responses to the target antigen of interest, particularly to viral envelope antigens. A preferred embodiment of such a composition comprises class I and class II epitopes in accordance with the invention. An alternative embodiment of such a composition comprises a class I

and/or class II epitope in accordance with the invention, along with a PADRE™ (Epimmune, San Diego, CA) molecule (described, for example, in U.S. Patent Number 5,736,142). Furthermore, any of these embodiments can be administered as a nucleic acid mediated modality.

5 The vaccine compositions of the invention may also be used in combination with antiviral drugs such as interferon- α .

For therapeutic or prophylactic immunization purposes, the peptides of the invention can also be expressed by viral or bacterial vectors. Examples of expression vectors include attenuated viral hosts, such as vaccinia or fowlpox. This approach
10 involves the use of vaccinia virus, for example, as a vector to express nucleotide sequences that encode the peptides of the invention. Upon introduction into an acutely or chronically infected host or into a non-infected host, the recombinant vaccinia virus expresses the immunogenic peptide, and thereby elicits a host CTL and/or HTL response. Vaccinia vectors and methods useful in immunization protocols are described in, *e.g.*,
15 U.S. Patent No. 4,722,848. Another vector is BCG (Bacille Calmette Guerin). BCG vectors are described in Stover *et al.*, *Nature* 351:456-460 (1991). A wide variety of other vectors useful for therapeutic administration or immunization of the peptides of the invention, *e.g.* adeno and adeno-associated virus vectors, retroviral vectors, *Salmonella typhi* vectors, detoxified anthrax toxin vectors, and the like, will be apparent to those
20 skilled in the art from the description herein.

Antigenic peptides are used to elicit a CTL and/or HTL response *ex vivo*, as well. The resulting CTL or HTL cells, can be used to treat chronic infections, or tumors in patients that do not respond to other conventional forms of therapy, or will not respond to a therapeutic vaccine peptide or nucleic acid in accordance with the invention. *Ex vivo*
25 CTL or HTL responses to a particular antigen (infectious or tumor-associated antigen) are induced by incubating in tissue culture the patient's, or genetically compatible, CTL or HTL precursor cells together with a source of antigen-presenting cells (APC), such as dendritic cells, and the appropriate immunogenic peptide. After an appropriate incubation time (typically about 14 weeks), in which the precursor cells are activated and
30 expanded into effector cells, the cells are infused back into the patient, where they will destroy (CTL) or facilitate destruction (HTL) of their specific target cell (an infected cell or a tumor cell). Transfected dendritic cells may also be used as antigen presenting cells. Alternatively, dendritic cells are transfected, *e.g.*, with a minigene construct in accordance

with the invention, in order to elicit immune responses. Minigenes will be discussed in greater detail in a following section.

Vaccine compositions may also be administered *in vivo* in combination with dendritic cell mobilization whereby loading of dendritic cells occurs *in vivo*.

5 DNA or RNA encoding one or more of the peptides of the invention can also be administered to a patient. This approach is described, for instance, in Wolff *et. al.*, *Science* 247:1465 (1990) as well as U.S. Patent Nos. 5,580,859; 5,589,466; 5,804,566; 5,739,118; 5,736,524; 5,679,647; WO 98/04720; and in more detail below. Examples of DNA-based delivery technologies include "naked DNA", facilitated (bupivacaine,
10 polymers, peptide-mediated) delivery, cationic lipid complexes, and particle-mediated ("gene gun") delivery.

Preferably, the following principles are utilized when selecting an array of epitopes for inclusion in a polyepitopic composition for use in a vaccine, or for selecting discrete epitopes to be included in a vaccine and/or to be encoded by nucleic acids such as
15 a minigene. Exemplary epitopes that may be utilized in a vaccine to treat or prevent HCV infection are set out in Tables XXVI-XXIX, and Table XXXII. It is preferred that each of the following principles are balanced in order to make the selection.

1.) Epitopes are selected which, upon administration, mimic immune responses that have been observed to be correlated with HCV clearance. For HLA Class I
20 this includes 3-4 epitopes that come from at least one antigen of HCV. For HLA Class II a similar rationale is employed; again 3-4 epitopes are selected from at least one HCV antigen (*see e.g.*, Rosenberg *et al.*, *Science* 278:1447-1450).

2.) Epitopes are selected that have the requisite binding affinity established to be correlated with immunogenicity: for HLA Class I an IC_{50} of 500 nM or less, or for
25 Class II an IC_{50} of 1000 nM or less.

3.) Sufficient supermotif bearing-peptides, or a sufficient array of allele-specific motif-bearing peptides, are selected to give broad population coverage. For example, it is preferable to have at least 80% population coverage. A Monte Carlo analysis, a statistical evaluation known in the art, can be employed to assess the breadth,
30 or redundancy of, population coverage.

4.) When selecting epitopes from cancer-related antigens it is often preferred to select analogs because the patient may have developed tolerance to the native epitope. When selecting epitopes for infectious disease-related antigens it is preferable to select either native or analoged epitopes. Of particular relevance for infectious disease vaccines

(but for cancer-related vaccines as well), are epitopes referred to as "nested epitopes." Nested epitopes occur where at least two epitopes overlap in a given peptide sequence. A peptide comprising "transcendent nested epitopes" is a peptide that has both HLA class I and HLA class II epitopes in it.

5 When providing nested epitopes, it is preferable to provide a sequence that has the greatest number of epitopes per provided sequence. Preferably, one avoids providing a peptide that is any longer than the amino terminus of the amino terminal epitope and the carboxyl terminus of the carboxyl terminal epitope in the peptide. When providing a longer peptide sequence, such as a sequence comprising nested epitopes, it is important to
10 screen the sequence in order to insure that it does not have pathological or other deleterious biological properties.

 5.) When creating a minigene, as disclosed in greater detail in the following section, an objective is to generate the smallest peptide possible that encompasses the epitopes of interest. The principles employed are similar, if not the same as those
15 employed when selecting a peptide comprising nested epitopes. Furthermore, upon determination of the nucleic acid sequence to be provided as a minigene, the peptide encoded thereby is analyzed to determine whether any "junctional epitopes" have been created. A junctional epitope is an actual binding epitope, as predicted, *e.g.*, by motif analysis, that only exists because two discrete peptide sequences are encoded directly
20 next to each other. Junctional epitopes are generally to be avoided because the recipient may generate an immune response to that non-native epitope. Of particular concern is a junctional epitope that is a "dominant epitope." A dominant epitope may lead to such a zealous response that immune responses to other epitopes are diminished or suppressed.

 Polyepitopic vaccine compositions may include epitopes from the core, S, E1,
25 NS1/E2, NS2, NS3, NS4, and NS5 domains of the HCV polyprotein. These regions encompass the following amino acid sequences using numbering relative to the prototype HCV-1 strain (Genbank accession number M62321; *see, e.g.*, US Patent Nos. 5,683,864 and 5,670,153): C domain (amino acids 1-120); S (amino acids 120-400); NS3 (amino acids 1050-1640); NS4 (amino acids 1640-2000); NS5 (amino acids 2000-3011); and
30 envelop proteins, E1 and E2/NS1, encompassing amino acids 192-750. Amino acids 750 to 1050 are designated as domain X as applied to the present invention. As appreciated by one of ordinary skill in the art, the designation of the amino acid range for each domain may diverge to some extent from that of HCV-1 depending on the strain of HCV.

One of ordinary skill in the art, when looking at an HCV polyprotein sequence, would readily be able to determine the domain boundaries.

Specific embodiments of the polyepitopic compositions of the present invention include a pharmaceutical composition comprising a pharmaceutically acceptable carrier and combination of motif-bearing peptides that are immunologically cross-reactive with peptides of HCV-1, wherein at least one of the peptides bears a motif of Table Ia, and further wherein the combination of motif-bearing peptides consists of: a) one or more peptides comprising at least 8 amino acids from an HCV C domain; b) one or more peptides comprising at least 8 amino acids of a further domain selected from the group consisting of: an S domain, an NS3 domain, an NS4 domain, or an NS5 domain, and; c) optionally, one or more motif-bearing peptides from one or more additional HCV domains with a *proviso* that an additional domain is not a further domain listed in "b". Preferably, such a pharmaceutical composition may additionally comprise one or more distinct HCV motif-bearing peptide(s) comprising at least 8 amino acids of an X domain or, alternatively, the composition may further comprise additional HCV motif-bearing peptide(s) that are from an envelope domain, the envelope domain peptide(s) consisting of one or more copies of a single HCV peptide comprising at least 8 amino acids of an envelope domain.

In another embodiment, the polyepitopic pharmaceutical composition may comprise a pharmaceutically acceptable carrier and combination of motif-bearing peptides that are immunologically cross-reactive with HCV-1 peptides, the peptides from multiple domains of HCV, wherein at least one of the peptides bears a motif of Table Ia, and wherein the combination of motif-bearing peptides consists essentially of: a) one or more peptides comprising at least 8 amino acids from a C domain; and, b) one or more peptides comprising at least 8 amino acids from an S, NS3, NS4, or NS5 domain, and, one HCV peptide comprising at least 8 amino acids of an envelope domain. Such a composition may further comprise one or more HCV motif-bearing peptides comprising at least 8 amino acids of an X domain.

Alternatively, a pharmaceutical composition of the invention may comprise: a) a pharmaceutically acceptable carrier; and, b) a combination of one or more motif-bearing peptides of at least 8 amino acids derived from one or more hepatitis C virus (HCV) domains, wherein said peptides are cross-reactive with peptides of HCV-1, with a *proviso* that the combination does not include a peptide of at least 8 amino acids from an HCV C domain, and wherein at least one of the peptides bears a motif of Table Ia, said domains

selected from the group consisting of: an S domain; an NS3 domain; an NS4 domain; an NS5 domain; and, an X domain. Such a composition may additionally comprise motif-bearing HCV envelope peptide(s) consisting of one or more copies of a single HCV peptide comprising at least 8 amino acids of an envelope domain.

5 Lastly, an embodiment of the invention may comprise a pharmaceutical composition comprising a pharmaceutically acceptable carrier and combination of two or more motif-bearing peptides from a single domain of an HCV-1 strain, said peptides immunologically cross-reactive with peptides of an HCV-1 antigen, wherein at least one of the peptides bears a motif of Table Ia, and the peptides are derived from HCV, and the
10 HCV domain is selected from the group consisting of: a C domain; an S domain; an NS3 domain; an NS4 domain; an NS5 domain; an X domain; or, an envelope domain from a single HCV strain, with a *proviso* that the envelope domain is other than a variable envelope domain.

In the embodiments set forth, "peptides immunologically cross-reactive with
15 HCV-1" refers to peptides that are bound by the same antibody; "derived from" refers to a fragment or subsequence and conservatively modified variants thereof.

IV.K.1. Minigene Vaccines

A growing body of experimental evidence demonstrates that a number of different
20 approaches are available which allow simultaneous delivery of multiple epitopes. Nucleic acids encoding the peptides of the invention are a particularly useful embodiment of the invention. Epitopes for inclusion in a minigene are preferably selected according to the guidelines set forth in the previous section. A preferred means of administering nucleic acids encoding the peptides of the invention uses minigene constructs encoding a
25 peptide comprising one or multiple epitopes of the invention. The use of multi-epitope minigenes is described below and in, *e.g.*, co-pending application U.S.S.N. 09/311,784; An, L. and Whitton, J. L., *J. Virol.* 71:2292, 1997; Thomson, S. A. *et al.*, *J. Immunol.* 157:822, 1996; Whitton, J. L. *et al.*, *J. Virol.* 67:348, 1993; Hanke, R. *et al.*, *Vaccine* 16:426, 1998. For example, a multi-epitope DNA plasmid encoding nine dominant HLA-
30 A*0201- and A11-restricted epitopes derived from the polymerase, envelope, and core proteins of HBV and human immunodeficiency virus (HIV), the PADRE™ universal helper T cell (HTL) epitope, and an endoplasmic reticulum-translocating signal sequence was engineered. Immunization of HLA transgenic mice with this plasmid construct

resulted in strong CTL induction responses against the nine epitopes tested, similar to those observed with a lipopeptide of known immunogenicity in humans, and significantly greater than immunization in oil-based adjuvants. Moreover, the immunogenicity of DNA-encoded epitopes *in vivo* correlated with the *in vitro* responses of specific CTL lines against target cells transfected with the DNA plasmid. Thus, these data show that the minigene served to both: 1.) generate a CTL response and 2.) that the induced CTLs recognized cells expressing the encoded epitopes. A similar approach may be used to develop minigenes encoding HCV epitopes.

For example, to create a DNA sequence encoding the selected epitopes (minigene) for expression in human cells, the amino acid sequences of the epitopes may be reverse translated. A human codon usage table can be used to guide the codon choice for each amino acid. These epitope-encoding DNA sequences may be directly adjoined, so that when translated, a continuous polypeptide sequence is created. To optimize expression and/or immunogenicity, additional elements can be incorporated into the minigene design. Examples of amino acid sequences that can be reverse translated and included in the minigene sequence include: HLA class I epitopes, HLA class II epitopes, a ubiquitination signal sequence, and/or an endoplasmic reticulum targeting signal. In addition, HLA presentation of CTL and HTL epitopes may be improved by including synthetic (*e.g.* poly-alanine) or naturally-occurring flanking sequences adjacent to the CTL or HTL epitopes; these larger peptides comprising the epitope(s) are within the scope of the invention.

The minigene sequence may be converted to DNA by assembling oligonucleotides that encode the plus and minus strands of the minigene. Overlapping oligonucleotides (30-100 bases long) may be synthesized, phosphorylated, purified and annealed under appropriate conditions using well known techniques. The ends of the oligonucleotides can be joined, for example, using T4 DNA ligase. This synthetic minigene, encoding the epitope polypeptide, can then be cloned into a desired expression vector.

Standard regulatory sequences well known to those of skill in the art are preferably included in the vector to ensure expression in the target cells. Several vector elements are desirable: a promoter with a down-stream cloning site for minigene insertion; a polyadenylation signal for efficient transcription termination; an *E. coli* origin of replication; and an *E. coli* selectable marker (*e.g.* ampicillin or kanamycin resistance). Numerous promoters can be used for this purpose, *e.g.*, the human cytomegalovirus

(hCMV) promoter. See, *e.g.*, U.S. Patent Nos. 5,580,859 and 5,589,466 for other suitable promoter sequences.

Additional vector modifications may be desired to optimize minigene expression and immunogenicity. In some cases, introns are required for efficient gene expression, and one or more synthetic or naturally-occurring introns could be incorporated into the transcribed region of the minigene. The inclusion of mRNA stabilization sequences and sequences for replication in mammalian cells may also be considered for increasing minigene expression.

Once an expression vector is selected, the minigene is cloned into the polylinker region downstream of the promoter. This plasmid is transformed into an appropriate *E. coli* strain, and DNA is prepared using standard techniques. The orientation and DNA sequence of the minigene, as well as all other elements included in the vector, are confirmed using restriction mapping and DNA sequence analysis. Bacterial cells harboring the correct plasmid can be stored as a master cell bank and a working cell bank.

In addition, immunostimulatory sequences (ISSs or CpGs) appear to play a role in the immunogenicity of DNA vaccines. These sequences may be included in the vector, outside the minigene coding sequence, if desired to enhance immunogenicity.

In some embodiments, a bi-cistronic expression vector which allows production of both the minigene-encoded epitopes and a second protein (included to enhance or decrease immunogenicity) can be used. Examples of proteins or polypeptides that could beneficially enhance the immune response if co-expressed include cytokines (*e.g.*, IL-2, IL-12, GM-CSF), cytokine-inducing molecules (*e.g.*, LeIF), costimulatory molecules, or for HTL responses, pan-DR binding proteins (PADRE™, Epimmune, San Diego, CA). Helper (HTL) epitopes can be joined to intracellular targeting signals and expressed separately from expressed CTL epitopes; this allows direction of the HTL epitopes to a cell compartment different than that of the CTL epitopes. If required, this could facilitate more efficient entry of HTL epitopes into the HLA class II pathway, thereby improving HTL induction. In contrast to HTL or CTL induction, specifically decreasing the immune response by co-expression of immunosuppressive molecules (*e.g.* TGF- β) may be beneficial in certain diseases.

Therapeutic quantities of plasmid DNA can be produced for example, by fermentation in *E. coli*, followed by purification. Aliquots from the working cell bank are used to inoculate growth medium, and grown to saturation in shaker flasks or a bioreactor

according to well known techniques. Plasmid DNA can be purified using standard bioseparation technologies such as solid phase anion-exchange resins supplied by QIAGEN, Inc. (Valencia, California). If required, supercoiled DNA can be isolated from the open circular and linear forms using gel electrophoresis or other methods.

5 Purified plasmid DNA can be prepared for injection using a variety of formulations. The simplest of these is reconstitution of lyophilized DNA in sterile phosphate-buffer saline (PBS). This approach, known as "naked DNA," is currently being used for intramuscular (IM) administration in clinical trials. To maximize the immunotherapeutic effects of minigene DNA vaccines, an alternative method for
10 formulating purified plasmid DNA may be desirable. A variety of methods have been described, and new techniques may become available. Cationic lipids can also be used in the formulation (see, *e.g.*, as described by WO 93/24640; Mannino & Gould-Fogerite, *BioTechniques* 6(7): 682 (1988); U.S. Pat No. 5,279,833; WO 91/06309; and Felgner, *et al.*, *Proc. Nat'l Acad. Sci. USA* 84:7413 (1987). In addition, glycolipids, fusogenic
15 liposomes, peptides and compounds referred to collectively as protective, interactive, non-condensing compounds (PINC) could also be complexed to purified plasmid DNA to influence variables such as stability, intramuscular dispersion, or trafficking to specific organs or cell types.

Target cell sensitization can be used as a functional assay for expression and HLA
20 class I presentation of minigene-encoded CTL epitopes. For example, the plasmid DNA is introduced into a mammalian cell line that is suitable as a target for standard CTL chromium release assays. The transfection method used will be dependent on the final formulation. Electroporation can be used for "naked" DNA, whereas cationic lipids allow direct *in vitro* transfection. A plasmid expressing green fluorescent protein (GFP) can be
25 co-transfected to allow enrichment of transfected cells using fluorescence activated cell sorting (FACS). These cells are then chromium-51 (^{51}Cr) labeled and used as target cells for epitope-specific CTL lines; cytolysis, detected by ^{51}Cr release, indicates both production of, and HLA presentation of, minigene-encoded CTL epitopes. Expression of HTL epitopes may be evaluated in an analogous manner using assays to assess HTL
30 activity.

In vivo immunogenicity is a second approach for functional testing of minigene DNA formulations. Transgenic mice expressing appropriate human HLA proteins are immunized with the DNA product. The dose and route of administration are formulation dependent (*e.g.*, IM for DNA in PBS, intraperitoneal (IP) for lipid-complexed DNA).

Twenty-one days after immunization, splenocytes are harvested and restimulated for 1 week in the presence of peptides encoding each epitope being tested. Thereafter, for CTL effector cells, assays are conducted for cytotoxicity of peptide-loaded, ^{51}Cr -labeled target cells using standard techniques. Lysis of target cells that were sensitized by HLA loaded with peptide epitopes, corresponding to minigene-encoded epitopes, demonstrates DNA vaccine function for *in vivo* induction of CTLs. Immunogenicity of HTL epitopes is evaluated in transgenic mice in an analogous manner.

Alternatively, the nucleic acids can be administered using ballistic delivery as described, for instance, in U.S. Patent No. 5,204,253. Using this technique, particles comprised solely of DNA are administered. In a further alternative embodiment, DNA can be adhered to particles, such as gold particles.

IV.K.2. Combinations of CTL Peptides with Helper Peptides

Vaccine compositions comprising the peptides of the present invention, or analogs thereof, which have immunostimulatory activity may be modified to provide desired attributes, such as improved serum half life, or to enhance immunogenicity.

For instance, the ability of the peptides to induce CTL activity can be enhanced by linking the peptide to a sequence which contains at least one epitope that is capable of inducing a T helper cell response. The use of T helper epitopes in conjunction with CTL epitopes to enhance immunogenicity is illustrated, for example, in co-pending U.S.S.N. 08/820360, U.S.S.N. 08/197,484, and U.S.S.N. 08/464,234.

Particularly preferred CTL epitope/HTL epitope conjugates are linked by a spacer molecule. The spacer is typically comprised of relatively small, neutral molecules, such as amino acids or amino acid mimetics, which are substantially uncharged under physiological conditions. The spacers are typically selected from, *e.g.*, Ala, Gly, or other neutral spacers of nonpolar amino acids or neutral polar amino acids. It will be understood that the optionally present spacer need not be comprised of the same residues and thus may be a hetero- or homo-oligomer. When present, the spacer will usually be at least one or two residues, more usually three to six residues. Alternatively, the CTL peptide may be linked to the T helper peptide without a spacer.

The CTL peptide epitope may be linked to the T helper peptide epitope either directly or via a spacer either at the amino or carboxy terminus of the CTL peptide. The amino terminus of either the immunogenic peptide or the T helper peptide may be acylated. The HTL peptide epitopes used in the invention can be modified in the same

manner as CTL peptides. For instance, they may be modified to include D-amino acids or be conjugated to other molecules such as lipids, proteins, sugars and the like. Exemplary T helper peptides include tetanus toxoid 830-843, influenza 307-319, and malarial circumsporozoite 382-398 and 378-398.

5 In certain embodiments, the T helper peptide is one that is recognized by T helper cells present in the majority of the population. This can be accomplished by selecting amino acid sequences that bind to many, most, or all of the HLA class II molecules. These are known as "loosely HLA-restricted" or "promiscuous" T helper sequences. Examples of amino acid sequences that are promiscuous include sequences from antigens
10 such as tetanus toxoid at positions 830-843 (QYIKANSKFIGITE), *Plasmodium falciparum* CS protein at positions 378-398 (DIEKKIAKMEKASSVFNVVNS), and Streptococcus 18kD protein at positions 116 (GAVDSILGGVATYGAA). Other examples include peptides bearing a DR 1-4-7 supermotif, or either of the DR3 motifs.

Alternatively, it is possible to prepare synthetic peptides capable of stimulating T
15 helper lymphocytes, in a loosely HLA-restricted fashion, using amino acid sequences not found in nature (*see, e.g.*, PCT publication WO 95/07707). These synthetic compounds called Pan-DR-binding epitopes (*e.g.*, PADRE™, Epimmune, Inc., San Diego, CA) are designed to most preferably bind most HLA-DR (human HLA class II) molecules. For instance, a pan-DR-binding epitope peptide having the formula: aKXVWANTLKAAa,
20 where "X" is either cyclohexylalanine, phenylalanine, or tyrosine, and a is either D-alanine or L-alanine, has been found to bind to most HLA-DR alleles, and to stimulate the response of T helper lymphocytes from most individuals, regardless of their HLA type. An alternative of a pan-DR binding epitope comprises all "L" natural amino acids and can be provided in the form of nucleic acids that encode the epitope.

25 HTL peptide epitopes can also be modified to alter their biological properties. For example, peptides comprising HTL epitopes can contain D-amino acids to increase their resistance to proteases and thus extend their serum half-life. Also, the epitope peptides of the invention can be conjugated to other molecules such as lipids, proteins or sugars, or any other synthetic compounds, to increase their biological activity. Specifically, the T
30 helper peptide can be conjugated to one or more palmitic acid chains at either the amino or carboxyl termini.

In some embodiments it may be desirable to include in the pharmaceutical compositions of the invention at least one component which primes cytotoxic T

lymphocytes. Lipids have been identified as agents capable of priming CTL *in vivo* against viral antigens. For example, palmitic acid residues can be attached to the ϵ - and α -amino groups of a lysine residue and then linked, *e.g.*, via one or more linking residues such as Gly, Gly-Gly-, Ser, Ser-Ser, or the like, to an immunogenic peptide. The
 5 lipidated peptide can then be administered either directly in a micelle or particle, incorporated into a liposome, or emulsified in an adjuvant, *e.g.*, incomplete Freund's adjuvant. In a preferred embodiment, a particularly effective immunogenic comprises palmitic acid attached to ϵ - and α - amino groups of Lys, which is attached via linkage, *e.g.*, Ser-Ser, to the amino terminus of the immunogenic peptide.

10 As another example of lipid priming of CTL responses, *E. coli* lipoproteins, such as tripalmitoyl-S-glycerylcysteinylserine (P₃CSS) can be used to prime virus specific CTL when covalently attached to an appropriate peptide. (*See, e.g., Deres, et al., Nature* 342:561, 1989). Peptides of the invention can be coupled to P₃CSS, for example, and the lipopeptide administered to an individual to specifically prime a CTL response to
 15 the target antigen. Moreover, because the induction of neutralizing antibodies can also be primed with P₃CSS-conjugated epitopes, two such compositions can be combined to more effectively elicit both humoral and cell-mediated responses to infection.

As noted herein, additional amino acids can be added to the termini of a peptide to provide for ease of linking peptides one to another, for coupling to a carrier support or
 20 larger peptide, for modifying the physical or chemical properties of the peptide or oligopeptide, or the like. Amino acids such as tyrosine, cysteine, lysine, glutamic or aspartic acid, or the like, can be introduced at the C- or N-terminus of the peptide or oligopeptide, particularly class I peptides. However, it is to be noted that modification at the carboxyl terminus of a CTL epitope may, in some cases, alter binding characteristics
 25 of the peptide. In addition, the peptide or oligopeptide sequences can differ from the natural sequence by being modified by terminal-NH₂ acylation, *e.g.*, by alkanoyl (C₁-C₂₀) or thioglycolyl acetylation, terminal-carboxyl amidation, *e.g.*, ammonia, methylamine, *etc.* In some instances these modifications may provide sites for linking to a support or other molecule.

30

IV.L. Administration of Vaccines for Therapeutic or Prophylactic Purposes

The peptides of the present invention and pharmaceutical and vaccine compositions of the invention are useful for administration to mammals, particularly

humans, to treat and/or prevent HCV infection. Vaccine compositions containing the peptides of the invention are administered to a patient infected with HCV or to an individual susceptible to, or otherwise at risk for, HCV infection to elicit an immune response against HCV antigens and thus enhance the patient's own immune response capabilities. In therapeutic applications, peptide and/or nucleic acid compositions are administered to a patient in an amount sufficient to elicit an effective CTL and/or HTL response to the virus antigen and to cure or at least partially arrest or slow symptoms and/or complications. An amount adequate to accomplish this is defined as "therapeutically effective dose." Amounts effective for this use will depend on, *e.g.*, the particular composition administered, the manner of administration, the stage and severity of the disease being treated, the weight and general state of health of the patient, and the judgment of the prescribing physician.

The vaccine compositions of the invention may also be used purely as prophylactic agents. Generally the dosage for an initial prophylactic immunization generally occurs in a unit dosage range where the lower value is about 1, 5, 50, 500, or 1000 μg and the higher value is about 10,000; 20,000; 30,000; or 50,000 μg . Dosage values for a human typically range from about 500 μg to about 50,000 μg per 70 kilogram patient. This is followed by boosting dosages of between about 1.0 μg to about 50,000 μg of peptide administered at defined intervals from about four weeks to six months after the initial administration of vaccine. The immunogenicity of the vaccine may be assessed by measuring the specific activity of CTL and HTL obtained from a sample of the patient's blood.

As noted above, peptides comprising CTL and/or HTL epitopes of the invention induce immune responses when presented by HLA molecules and contacted with a CTL or HTL specific for an epitope comprised by the peptide. The manner in which the peptide is contacted with the CTL or HTL is not critical to the invention. For instance, the peptide can be contacted with the CTL or HTL either *in vivo* or *in vitro*. If the contacting occurs *in vivo*, the peptide itself can be administered to the patient, or other vehicles, *e.g.*, DNA vectors encoding one or more peptides, viral vectors encoding the peptide(s), liposomes and the like, can be used, as described herein.

For pharmaceutical compositions, the immunogenic peptides of the invention, or DNA encoding them, are generally administered to an individual already infected with HCV. The peptides or DNA encoding them can be administered individually or as fusions of one or more peptide sequences. Those in the incubation phase or the acute

phase of infection can be treated with the immunogenic peptides separately or in conjunction with other treatments, as appropriate.

For therapeutic use, administration should generally begin at the first diagnosis of HCV infection. This is followed by boosting doses until at least symptoms are
5 substantially abated and for a period thereafter. In chronic infection, loading doses followed by boosting doses may be required.

Treatment of an infected individual with the compositions of the invention may hasten resolution of the infection in acutely infected individuals. For those individuals susceptible (or predisposed) to developing chronic infection, the compositions are
10 particularly useful in methods for preventing the evolution from acute to chronic infection. Where susceptible individuals are identified prior to or during infection, the composition can be targeted to them, thus minimizing the need for administration to a larger population.

The peptide or other compositions used for the treatment or prophylaxis of HCV
15 infection can be used, *e.g.*, in persons who have not manifested symptoms of disease but who act as a disease vector. In this context, it is generally important to provide an amount of the peptide epitope delivered by a mode of administration sufficient to effectively stimulate a cytotoxic T cell response; compositions which stimulate helper T cell responses can also be given in accordance with this embodiment of the invention.

20 The dosage for an initial therapeutic immunization generally occurs in a unit dosage range where the lower value is about 1, 5, 50, 500, or 1000 μg and the higher value is about 10,000; 20,000; 30,000; or 50,000 μg . Dosage values for a human typically range from about 500 μg to about 50,000 μg per 70 kilogram patient. Boosting dosages of between about 1.0 μg to about 50000 μg of peptide pursuant to a boosting
25 regimen over weeks to months may be administered depending upon the patient's response and condition as determined by measuring the specific activity of CTL and HTL obtained from the patient's blood. The peptides and compositions of the present invention may be employed in serious disease states, that is, life-threatening or potentially life threatening situations. In such cases, as a result of the minimal amounts of
30 extraneous substances and the relative nontoxic nature of the peptides in preferred compositions of the invention, it is possible and may be felt desirable by the treating physician to administer substantial excesses of these peptide compositions relative to these stated dosage amounts.

Thus, for treatment of chronic infection, a representative dose is in the range disclosed above, namely where the lower value is about 1, 5, 50, 500, or 1000 μg and the higher value is about 10,000; 20,000; 30,000; or 50,000 μg , preferably from about 500 μg to about 50,000 μg per 70 kilogram patient. Initial doses followed by boosting doses at established intervals, *e.g.*, from four weeks to six months, may be required, possibly for a prolonged period of time to effectively immunize an individual. In the case of chronic infection, administration should continue until at least clinical symptoms or laboratory tests indicate that the viral infection has been eliminated or substantially abated and for a period thereafter. The dosages, routes of administration, and dose schedules are adjusted in accordance with methodologies known in the art.

The pharmaceutical compositions for therapeutic treatment are intended for parenteral, topical, oral, intrathecal, or local administration. Preferably, the pharmaceutical compositions are administered parentally, *e.g.*, intravenously, subcutaneously, intradermally, or intramuscularly. Thus, the invention provides compositions for parenteral administration which comprise a solution of the immunogenic peptides dissolved or suspended in an acceptable carrier, preferably an aqueous carrier. A variety of aqueous carriers may be used, *e.g.*, water, buffered water, 0.8% saline, 0.3% glycine, hyaluronic acid and the like. These compositions may be sterilized by conventional, well known sterilization techniques, or may be sterile filtered. The resulting aqueous solutions may be packaged for use as is, or lyophilized, the lyophilized preparation being combined with a sterile solution prior to administration. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions, such as pH-adjusting and buffering agents, tonicity adjusting agents, wetting agents, preservatives, and the like, for example, sodium acetate, sodium lactate, sodium chloride, potassium chloride, calcium chloride, sorbitan monolaurate, triethanolamine oleate, *etc.*

The concentration of peptides of the invention in the pharmaceutical formulations can vary widely, *i.e.*, from less than about 0.1%, usually at or at least about 2% to as much as 20% to 50% or more by weight, and will be selected primarily by fluid volumes, viscosities, *etc.*, in accordance with the particular mode of administration selected.

A human unit dose form of the peptide composition is typically included in a pharmaceutical composition that comprises a human unit dose of an acceptable carrier, preferably an aqueous carrier, and is administered in a volume of fluid that is known by those of skill in the art to be used for administration of such compositions to humans (*see*,

e.g., Remington's Pharmaceutical Sciences, 17th Edition, A. Gennaro, Editor, Mack Publishing Co., Easton, Pennsylvania, 1985).

The peptides of the invention may also be administered via liposomes, which serve to target the peptides to a particular tissue, such as lymphoid tissue, or to target selectively to infected cells, as well as to increase the half-life of the peptide composition. Liposomes include emulsions, foams, micelles, insoluble monolayers, liquid crystals, phospholipid dispersions, lamellar layers and the like. In these preparations, the peptide to be delivered is incorporated as part of a liposome, alone or in conjunction with a molecule which binds to a receptor prevalent among lymphoid cells, such as monoclonal antibodies which bind to the CD45 antigen, or with other therapeutic or immunogenic compositions. Thus, liposomes either filled or decorated with a desired peptide of the invention can be directed to the site of lymphoid cells, where the liposomes then deliver the peptide compositions. Liposomes for use in accordance with the invention are formed from standard vesicle-forming lipids, which generally include neutral and negatively charged phospholipids and a sterol, such as cholesterol. The selection of lipids is generally guided by consideration of, *e.g.*, liposome size, acid lability and stability of the liposomes in the blood stream. A variety of methods are available for preparing liposomes, as described in, *e.g.*, Szoka, *et al.*, *Ann. Rev. Biophys. Bioeng.* 9:467 (1980), and U.S. Patent Nos. 4,235,871, 4,501,728, 4,837,028, and 5,019,369.

For targeting cells of the immune system, a ligand to be incorporated into the liposome can include, *e.g.*, antibodies or fragments thereof specific for cell surface determinants of the desired immune system cells. A liposome suspension containing a peptide may be administered intravenously, locally, topically, *etc.* in a dose which varies according to, *inter alia*, the manner of administration, the peptide being delivered, and the stage of the disease being treated.

For solid compositions, conventional nontoxic solid carriers may be used which include, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharin, talcum, cellulose, glucose, sucrose, magnesium carbonate, and the like. For oral administration, a pharmaceutically acceptable nontoxic composition is formed by incorporating any of the normally employed excipients, such as those carriers previously listed, and generally 10-95% of active ingredient, that is, one or more peptides of the invention, and more preferably at a concentration of 25%-75%.

For aerosol administration, the immunogenic peptides are preferably supplied in finely divided form along with a surfactant and propellant. Typical percentages of

peptides are 0.01%-20% by weight, preferably 1%-10%. The surfactant must, of course, be nontoxic, and preferably soluble in the propellant. Representative of such agents are the esters or partial esters of fatty acids containing from 6 to 22 carbon atoms, such as caproic, octanoic, lauric, palmitic, stearic, linoleic, linolenic, olesteric and oleic acids with an aliphatic polyhydric alcohol or its cyclic anhydride. Mixed esters, such as mixed or natural glycerides may be employed. The surfactant may constitute 0.1%-20% by weight of the composition, preferably 0.25-5%. The balance of the composition is ordinarily propellant. A carrier can also be included, as desired, as with, *e.g.*, lecithin for intranasal delivery.

IV.M. Kits

The peptide and nucleic acid compositions of this invention can be provided in kit form together with instructions for vaccine administration. Typically the kit would include desired peptide compositions in a container, preferably in unit dosage form and instructions for administration. An alternative kit would include a minigene construct with desired nucleic acids of the invention in a container, preferably in unit dosage form together with instructions for administration. Lymphokines such as IL-2 or IL-12 may also be included in the kit. Other kit components that may also be desirable include, for example, a sterile syringe, booster dosages, and other desired excipients.

The invention will be described in greater detail by way of specific examples. The following examples are offered for illustrative purposes, and are not intended to limit the invention in any manner. Those of skill in the art will readily recognize a variety of non-critical parameters that can be changed or modified to yield alternative embodiments in accordance with the invention.

V. EXAMPLES

As in many viral diseases, there is evidence that clearance of HCV is mediated by CTL. In a study of primary HCV infection in six chimpanzees, four progressed to chronic infection (Cooper *et al.*, abstract, 19th US-Japan Hepatitis Joint Panel Meeting, January 27-29, 1998). It was found that these four animals showed either no CTL response or a very narrowly focused response during early infection. In contrast, in the remaining two animals that resolved the infection, a broad CTL response was observed against multiple HCV proteins, some of which were conserved. Weiner *et al.* (*Proc. Natl. Acad. Sci. USA* 92:2755-2759, 1995) demonstrated that viral escape, in which the

epitopes presented to PATR class I molecules mutated, was linked with a progression toward chronic infection. These data show a role for the CTL in directing the course of HCV disease, and in shaping the genetic composition of HCV species in the persistently infected host.

5 In work in humans, Koziel and co-workers have established the presence of HCV-specific CTL in liver infiltrates from patients with chronic HCV infection (Koziel *et al.*, *J. Immunol.* 149:3339, 1992; and Koziel *et al.*, *J. Virol.* 67:7522, 1993), and have also identified a number of CTL epitopes recognized in the context of several different HLA class I molecules. Other investigators have shown that HCV-specific CTL can be
10 detected in the peripheral blood of patients with chronic hepatitis C (Cerny *et al.*, *J. Clin. Invest.* 95:521, 1995; Cerny *et al.*, *Curr. Topics in Micro. and Immunol.* 189:169, 1994; Cerny *et al.*, Abst. 2nd International Meeting on Hepatitis C and Related Viruses; La Jolla, CA, 1994; Battegay *et al.*, Abst. 2nd International Meeting on Hepatitis C and Related Viruses; La Jolla, CA, 1994; Shirai *et al.*, *J. Virol.* 68:3334, 1994; Shirai *et al.*, *J.*
15 *Immunol.* 154:2733, 1995; Battegay *et al.*, *J. Virol.* 69:2462, 1995). In addition, escape variants have been demonstrated in patients chronically infected with HCV (Chang *et al.*, *J. Clin. Invest.* 100:2376-2385, 1997; Tsai *et al.*, *Gastroenterology* 115:954-966, 1998).

The magnitude of the CTL responses observed in HCV-infected patients is, in general, higher than those observed in the case of chronic hepatitis B infection,
20 suggesting that there is less impairment of specific T cell immunity than with HBV infection. The magnitude of CTL responses in HCV patients is, however, lower than those observed in HBV infected individuals who successfully cleared HBV infection. These results support the understanding that HCV infected patients are capable of responding to active immunotherapy, and suggest that potentiation and increasing of T
25 cell responses to HCV may be of use in therapy and prevention of chronic HCV infection (Prince, A. M. *FEMS Micro. Rev.* 14:273, 1994).

Several groups have analyzed the potential role of HCV-specific CTL responses in disease resistance and pathogenesis. In some studies no correlation was found between CTL viremia and CTL precursor frequency for individual HCV epitopes (Rehermann *et al.*, *J. Clin. Invest.* 98:1432-1440, 1996; Wong *et al.*, *J. Immunol.* 160:1479-1488, 1998).
30 In other studies, however, it was shown that a clear correlation existed between levels of HCV infection and CTL responses, provided that the global response against multiple CTL epitopes was considered (Rehermann *et al.*, *J. Virol.* 70:7092-7102, 1996). These data represent a strong rationale for development of vaccine constructs capable of

inducing vigorous CTL responses directed against a multiplicity of conserved HCV-derived epitopes.

Koziel and colleagues have demonstrated the presence of HCV-specific CTLs, as well as T helper cell responses, in exposed but seronegative individuals (Koziel *et al.*, *J. Infect. Diseases* 176:859-866, 1997). In addition, HCV-specific CTLs have been detected in healthy, seronegative family members of chronically HCV-infected patients, indicating that a protective immunity is established in absence of a detectable infection (Bronowicki *et al.*, *J. Infect. Dis.* 176:518-522, 1997; Scognamiglio *et al.*, in preparation).

Experimental evidence also indicates that HTL epitopes play an important role in immune reactivity and defenses against HCV infection (Missale *et al.*, *J. Clin. Invest.* 98:706-714, 1996). Diepolder *et al.* (in *Lancet* 346:1006, 1995) have shown that a region of the NS3 gene (NS3 1007-1534) is recognized by patients who clear acute HCV infection, but is not seen by patients who develop chronic infection. Subsequent studies have shown that this particular region contained a highly cross-reactive HTL epitope (NS3 1248-1261), which binds with good affinity to 10 of 13 DR molecules tested, and is highly conserved in 30/33 different HCV isolates considered (Diepolder *et al.*, *J. Virol.* 71:6011-6019, 1997). These data suggest that directing HTL responses to this type of epitope (rather than to less cross-reactive and/or highly variable ones) will be of therapeutic and prophylactic benefit and strongly argue for inclusion of this and other epitopes with similar characteristics in HCV vaccine constructs.

The following examples illustrate identification, selection, and use of immunogenic Class I and Class II peptide epitopes for inclusion in vaccine compositions.

Example 1: HLA Class I and Class II Binding Assays

The following example of peptide binding to HLA molecules demonstrates quantification of binding affinities of HLA class I and class II peptides. Binding assays can be performed with peptides that are either motif-bearing or not motif-bearing.

Epstein-Barr virus (EBV)-transformed homozygous cell lines, fibroblasts, CIR, or 721.22 transfectants were used as sources of HLA class I molecules. These cells were maintained *in vitro* by culture in RPMI 1640 medium supplemented with 2mM L-glutamine (GIBCO, Grand Island, NY), 50µM 2-ME, 100µg/ml of streptomycin, 100U/ml of penicillin (Irvine Scientific) and 10% heat-inactivated FCS (Irvine Scientific, Santa Ana, CA). Cells were grown in 225-cm² tissue culture flasks or, for large-scale

cultures, in roller bottle apparatuses. The specific cell lines routinely used for purification of MHC class I and class II molecules are listed in Table XXIV.

Cell lysates were prepared and HLA molecules purified in accordance with disclosed protocols (Sidney *et al.*, *Current Protocols in Immunology* 18.3.1 (1998); Sidney, *et al.*, *J. Immunol.* 154:247 (1995); Sette, *et al.*, *Mol. Immunol.* 31:813 (1994)). Briefly, cells were lysed at a concentration of 10^8 cells/ml in 50 mM Tris-HCl, pH 8.5, containing 1% Nonidet P-40 (Fluka Biochemika, Buchs, Switzerland), 150 mM NaCl, 5 mM EDTA, and 2 mM PMSF. Lysates were cleared of debris and nuclei by centrifugation at $15,000 \times g$ for 30min.

HLA molecules were purified from lysates by affinity chromatography. Lysates prepared as above were passed twice through two pre-columns of inactivated Sepharose CL4-B and protein A-Sepharose. Next, the lysate was passed over a column of Sepharose CL-4B beads coupled to an appropriate antibody. The antibodies used for the extraction of HLA from cell lysates are listed in Table XXV. The anti-HLA column was then washed with 10-column volumes of 10mM Tris-HCL, pH 8.0, in 1% NP-40, PBS, 2-column volumes of PBS, and 2-column volumes of PBS containing 0.4% n-octylglucoside. Finally, MHC molecules were eluted with 50mM diethylamine in 0.15M NaCl containing 0.4% n-octylglucoside, pH 11.5. A 1/25 volume of 2.0M Tris, pH 6.8, was added to the eluate to reduce the pH to ~8.0. Eluates were then be concentrated by centrifugation in Centriprep 30 concentrators at 2000 rpm (Amicon, Beverly, MA). Protein content was evaluated by a BCA protein assay (Pierce Chemical Co., Rockford, IL) and confirmed by SDS-PAGE.

A detailed description of the protocol utilized to measure the binding of peptides to Class I and Class II MHC has been published (Sette *et al.*, *Mol. Immunol.* 31:813, 1994; Sidney *et al.*, in *Current Protocols in Immunology*, Margulies, Ed., John Wiley & Sons, New York, Section 18.3, 1998). Briefly, purified MHC molecules (5 to 500nM) were incubated with various unlabeled peptide inhibitors and 1-10nM 125 I-radiolabeled probe peptides for 48h in PBS containing 0.05% Nonidet P-40 (NP40) (or 20% w/v digitonin for H-2 IA assays) in the presence of a protease inhibitor cocktail. The final concentrations of protease inhibitors (each from CalBioChem, La Jolla, CA) were 1 mM PMSF, 1.3 nM 1.10 phenanthroline, 73 μ M pepstatin A, 8mM EDTA, 6mM N-ethylmaleimide (for Class II assays), and 200 μ M N alpha-p-tosyl-L-lysine chloromethyl ketone (TLCK). All assays were performed at pH 7.0 with the exception of DRB1*0301.

which was performed at pH 4.5, and DRB1*1601 (DR2w21 β ₁) and DRB4*0101 (DRw53), which were performed at pH 5.0. pH was adjusted as described elsewhere (see Sidney *et al.*, in *Current Protocols in Immunology*, Margulies, Ed., John Wiley & Sons, New York, Section 18.3, 1998).

5 Following incubation, MHC-peptide complexes were separated from free peptide by gel filtration on 7.8 mm x 15 cm TSK200 columns (TosoHaas 16215, Montgomeryville, PA), eluted at 1.2 mls/min with PBS pH 6.5 containing 0.5% NP40 and 0.1% NaN₃. Because the large size of the radiolabeled peptide used for the DRB1*1501 (DR2w2 β ₁) assay makes separation of bound from unbound peaks more difficult under
10 these conditions, all DRB1*1501 (DR2w2 β ₁) assays were performed using a 7.8mm x 30cm TSK2000 column eluted at 0.6 mls/min. The eluate from the TSK columns was passed through a Beckman 170 radioisotope detector, and radioactivity was plotted and integrated using a Hewlett-Packard 3396A integrator, and the fraction of peptide bound was determined.

15 Radiolabeled peptides were iodinated using the chloramine-T method. Representative radiolabeled probe peptides utilized in each assay, and its assay specific IC₅₀ nM, are summarized in Tables IV and V. Typically, in preliminary experiments, each MHC preparation was titrated in the presence of fixed amounts of radiolabeled peptides to determine the concentration of HLA molecules necessary to bind 10-20% of
20 the total radioactivity. All subsequent inhibition and direct binding assays were performed using these HLA concentrations.

 Since under these conditions [label]<[HLA] and IC₅₀≥[HLA], the measured IC₅₀ values are reasonable approximations of the true K_D values. Peptide inhibitors are typically tested at concentrations ranging from 120 μ g/ml to 1.2 ng/ml, and are tested in
25 two to four completely independent experiments. To allow comparison of the data obtained in different experiments, a relative binding figure is calculated for each peptide by dividing the IC₅₀ of a positive control for inhibition by the IC₅₀ for each tested peptide (typically unlabeled versions of the radiolabeled probe peptide). For database purposes, and inter-experiment comparisons, relative binding values are compiled. These values
30 can subsequently be converted back into IC₅₀ nM values by dividing the IC₅₀ nM of the positive controls for inhibition by the relative binding of the peptide of interest. This method of data compilation has proven to be the most accurate and consistent for

comparing peptides that have been tested on different days, or with different lots of purified MHC.

Because the antibody used for HLA-DR purification (LB3.1) is α -chain specific, β_1 molecules are not separated from β_3 (and/or β_4 and β_5) molecules. The β_1 specificity of the binding assay is obvious in the cases of DRB1*0101 (DR1), DRB1*0802 (DR8w2), and DRB1*0803 (DR8w3), where no β_3 is expressed. It has also been demonstrated for DRB1*0301 (DR3) and DRB3*0101 (DR52a), DRB1*0401 (DR4w4), DRB1*0404 (DR4w14), DRB1*0405 (DR4w15), DRB1*1101 (DR5), DRB1*1201 (DR5w12), DRB1*1302 (DR6w19) and DRB1*0701 (DR7). The problem of β chain specificity for DRB1*1501 (DR2w2 β_1), DRB5*0101 (DR2w2 β_2), DRB1*1601 (DR2w21 β_1), DRB5*0201 (DR51Dw21), and DRB4*0101 (DRw53) assays is circumvented by the use of fibroblasts. Development and validation of assays with regard to DR β molecule specificity have been described previously (*see, e.g., Southwood et al., J. Immunol.* 160:3363-3373, 1998).

Binding assays as outlined above may be used to analyze supermotif and/or motif-bearing epitopes as, for example, described in Example 2.

Example 2. Identification of Conserved HLA Supermotif- and Motif-Bearing CTL Candidate Epitopes

Vaccine compositions of the invention may include multiple epitopes that comprise multiple HLA supermotifs or motifs to achieve broad population coverage. This example illustrates the identification of supermotif- and motif-bearing epitopes for the inclusion in such a vaccine composition. Calculation of population coverage was performed using the strategy described below.

Computer searches and algorithms for identification of supermotif and/or motif-bearing epitopes

Computer searches for epitopes bearing HLA Class I or Class II supermotifs or motifs were performed as follows. All translated HCV isolate sequences were analyzed using a text string search software program, *e.g.,* MotifSearch 1.4 (D. Brown, San Diego) to identify potential peptide sequences containing appropriate HLA binding motifs; alternative programs are readily produced in accordance with information in the art in view of the motif/supermotif disclosure herein. Furthermore, such calculations can be

made mentally. Identified A2-, A3-, and DR-supermotif sequences were scored using polynomial algorithms to predict their capacity to bind to specific HLA-Class I or Class II molecules. These polynomial algorithms take into account both extended and refined motifs (that is, to account for the impact of different amino acids at different positions),
 5 and are essentially based on the premise that the overall affinity (or ΔG) of peptide-HLA molecule interactions can be approximated as a linear polynomial function of the type:

$$“\Delta G” = a_{1i} \times a_{2i} \times a_{3i} \dots \times a_{ni}$$

where a_{ji} is a coefficient which represents the effect of the presence of a given amino acid (j) at a given position (i) along the sequence of a peptide of n amino acids. The crucial
 10 assumption of this method is that the effects at each position are essentially independent of each other (i.e., independent binding of individual side-chains). When residue j occurs at position i in the peptide, it is assumed to contribute a constant amount j_i to the free energy of binding of the peptide irrespective of the sequence of the rest of the peptide. This assumption is justified by studies from our laboratories that demonstrated that
 15 peptides are bound to MHC and recognized by T cells in essentially an extended conformation (data omitted herein).

The method of derivation of specific algorithm coefficients has been described in Gulukota *et al.*, *J. Mol. Biol.* 267:1258-126, 1997; (see also Sidney *et al.*, *Human Immunol.* 45:79-93, 1996; and Southwood *et al.*, *J. Immunol.* 160:3363-3373, 1998).
 20 Briefly, for all i positions, anchor and non-anchor alike, the geometric mean of the average relative binding (ARB) of all peptides carrying j is calculated relative to the remainder of the group, and used as the estimate of j_i . For Class II peptides, if multiple alignments are possible, only the highest scoring alignment is utilized, following an iterative procedure. To calculate an algorithm score of a given peptide in a test set, the
 25 ARB values corresponding to the sequence of the peptide are multiplied. If this product exceeds a chosen threshold, the peptide is predicted to bind. Appropriate thresholds are chosen as a function of the degree of stringency of prediction desired.

Selection of HLA-A2 supertype cross-reactive peptides

30 Complete polyprotein sequences from fourteen HCV isolates were aligned, then scanned, utilizing motif identification software, to identify conserved 9- and 10-mer sequences containing the HLA-A2-supermotif main anchor specificity.

A total of 231 conserved, HLA-A2 supermotif-positive sequences were identified. These peptides were then evaluated for the presence of A*0201 preferred secondary anchor residues using A*0201-specific polynomial algorithms. A total of 67 conserved, motif-bearing and algorithm-positive sequences were identified.

5 Fifty of these conserved, motif-containing 9- and 10-mer peptides were tested for their capacity to bind to purified HLA-A*0201 molecules *in vitro* (HLA-A*0201 is considered a prototype A2 supertype molecule). Sixteen peptides bound A*0201 with IC_{50} values ≤ 500 nM; 4 with high binding affinities (IC_{50} values ≤ 50 nM) and 12 with intermediate binding affinities, in the 50-500 nM range (Table XXVI).

10 These 16 peptides were subsequently tested for the capacity to bind to additional A2-supertype molecules (A*0202, A*0203, A*0206, and A*6802). As shown in Table XXVI, most of these peptides were found to be A2-supertype cross-reactive binders. More specifically, 12/16 (75%) peptides bound at least three of the five A2-supertype molecules tested.

15 *Selection of HLA-A3 supermotif-bearing epitopes*

The sequences from the same fourteen known HCV isolates scanned above were also examined for the presence of conserved peptides with the HLA-A3-supermotif primary anchors. A total of 71 conserved 9- or 10-mer motif containing sequences were
20 identified. Further analysis using the A03 and A11 algorithms (see, e.g., Gulukota et al, *J. Mol. Biol.* 267:1258-1267, 1997 and Sidney et al, *Human Immunol.* 45:79-93, 1996) identified 39 sequences that scored high in either or both algorithms. Twenty seven of the 39 peptides were synthesized and tested for binding to HLA-A*03 and HLA-A*11, the two most prevalent A3-supertype molecules. Fifteen peptides were identified which
25 bound A3 and/or A11 with binding affinities of ≤ 500 nM (Table XXVII). These peptides were then tested for binding cross-reactivity to the other common A3-supertype alleles (A*3101, A*3301, and A*6801). Seven of the 15 peptides bound at least three of the five HLA-A3-supertype molecules tested.

In the course of an independent series of experiments (Kubo *et al.*, *J. Immunol.*
30 152:3913-3924, 1994), one peptide, HCV NS3 1262, not identified by the selection criteria utilized above because it does not have the A3-supermotif main anchor specificity, was determined to be cross-reactive in the A3-supertype, binding A*03, A*11, and A*6801. It is also shown in Table XXVII. Interestingly, this peptide

represents a single residue N-terminal truncation of peptide 1073.14, which is also shown in Table XXVII.

In summary, 8 peptides that bind 3 or more A3-supertype molecules derived from conserved regions of the HCV genome were identified.

5

Selection of HLA-B7 supermotif bearing epitopes

When the same fourteen HCV isolates were also analyzed for the presence of conserved 9- or 10-mer peptides with the HLA-B7-supermotif, 35 sequences were identified. The corresponding peptides were synthesized and tested for binding to HLA-B*0702, the most common B7-supertype allele (*i.e.*, the prototype B7 supertype allele). Thirteen peptides bound B*0702 with IC_{50} of ≤ 500 nM (Table XXVIIIa). These 13 peptides were then tested for binding to other common B7-supertype molecules (B*3501, B*51, B*5301, and B*5401). As shown in Table XXVIIIa, only 1 peptide (Core 169) was capable of binding to three or more of the five B7-supertype alleles tested.

15

To identify additional B7-supertype epitopes, further studies were undertaken. The protein sequences from the fourteen HCV isolates utilized above were again examined to identify conserved, motif-containing 8- and 11-mers. The isolates were also examined for 9- and 10-mer sequences allowing for lower conservancy (51%-78%). These analyses identified twenty-five 8-mers, sixteen 11-mers, and thirty-five 9- and 10-mers. These peptides were synthesized and tested for binding to B*0702. Thirteen peptides bound with high or intermediate affinity to B*0702 ($IC_{50} \leq 500$ nM) (Table XXVIIIb). These peptides were additionally screened for binding to other B7-supertype molecules. Only one cross-reactive binder, the NS3 1378 8-mer (peptide 29.0035/1260.04), was identified (Table XXVIIIb).

20

In summary, a total of two cross-reactive B7-supertype binders were identified (Core 169 and NS3 1378).

Selection of A1 and A24 motif-bearing epitopes

To further increase population coverage, HLA-A1 and -A24 epitopes can also be incorporated into potential vaccine constructs.

30

In a previous analysis, two A1 and three A24 binders, 100% conserved among four strains of HCV, were identified (Wentworth *et al.*, *Int. Immunol.* 8:651-659, 1996). An analysis of the protein sequence data from the fourteen HCV strains utilized above

demonstrated that these peptides were >79% conserved, and also identified an additional eleven A1- and twenty five A24-motif-containing conserved sequences (see Table XXIXA and B). Testing for binding to the appropriate HLA molecule (*i.e.*, A1 or A24) was completed for eight of the additional eleven A1 peptides, and seven of the additional
 5 twenty five A24 peptides. Overall, as shown in Table XXIX, four A1-motif peptides (A) and three A24-motif peptides (B) have been found with binding capacities of 500 nM or less for the appropriate allele-specific HLA molecule.

Analysis of the HLA-A2 and A3 supermotif-bearing epitopes identified above revealed that in 13/14 cases, peptides binding the supertype prototype HLA molecule (*i.e.*
 10 A*0201 for the A2 supertype, and A*0301 for the A3 supertype) with an IC₅₀ of less than 100nM were cross-reactive and recognized by HCV-infected patients as described in Example 3, which follows. Based on these observations, two A1 peptides and one A24 peptide epitopes were also selected as candidates for inclusion in vaccine compositions; these peptides bind the appropriate HLA molecule with an IC₅₀ of less than 100nM.

Example 3: Confirmation of Immunogenicity

*Evaluation of A*0201 immunogenicity*

It has been shown that CTL induced in A*0201/K^b transgenic mice exhibit specificity similar to CTL induced in the human system (*see, e.g.*, Vitiello *et al.*, *J. Exp.*
 20 *Med.* 173:1007-1015, 1991; Wentworth *et al.*, *Eur. J. Immunol.* 26:97-101, 1996). Accordingly, these mice were used to evaluate the immunogenicity of the twelve conserved A2-supertype cross-reactive peptides identified in Example 2 above.

CTL induction in transgenic mice following peptide immunization has been described (Vitiello *et al.*, *J. Exp. Med.* 173:1007-1015, 1991; Alexander *et al.*; *J.*
 25 *Immunol.* 159:4753-4761, 1997). In these studies, mice were injected subcutaneously at the base of the tail with each peptide (50 µg/mouse) emulsified in IFA in the presence of an excess of an IA^b-restricted helper peptide (140 µg/mouse) (HBV core 128-140, Sette *et al.*, *J. Immunol.* 153:5586-5592, 1994). Eleven days after injection, splenocytes were incubated in the presence of peptide-loaded syngenic LPS blasts. After six days, cultures
 30 were assayed for cytotoxic activity using peptide-pulsed targets. The data, summarized in Table XXX, indicate that 7 of the 12 peptides (58%) were capable of inducing primary CTL responses in A*0201/K^b transgenic mice. (For these studies, a peptide was

considered positive if it induced CTL (L.U. $30/10^6$ cells ≥ 2 in at least two transgenic animals (Wentworth *et al.*, *Eur. J. Immunol.* 26:97-101, 1996).

The conserved, cross reactive candidate CTL epitopes were also tested for recognition *in vitro* by PBMCs obtained from HCV-infected patients. Briefly, PBMC
 5 from patients infected with HCV were cultured in the presence of 10 $\mu\text{g}/\text{ml}$ of synthetic peptide. After 7 and 14 days, the cultures were restimulated with peptide. The cultures were assayed for cytolytic activity on day 21 using target cells pulsed with the specific peptide in a standard four hour ^{51}Cr release assay. The data are summarized in Table XXX. As shown, all 12 peptides are CTL epitopes recognized by PBMC from HCV-
 10 infected patients. From the data in Table XXX, it is interesting to note that HLA transgenics did not fully reveal the immunogenicity of some peptides that were positive in recall responses. This apparent discrepancy may reflect differences in the route of immunization utilized (*e.g.*, natural infection versus peptide immunization), or CTL repertoire.

15

*Evaluation of A*03/A11 immunogenicity*

The immunogenicity of six of the eight A3-supertype cross-reactive peptides identified in Example 2 above was evaluated in HLA-A11/K^b transgenic mice, using the protocol described above for HLA-A2 transgenic mice (Alexander *et al.*, *J. Immunol.*
 20 159:4753-4761, 1997). Five of these six peptides were able to induce primary CTL responses (Table XXXI).

All eight peptides were also studied by collaborators using PBMC cultures from HCV infected patients and contacts of such patients. This data is also summarized in Table XXXI. Briefly, all eight peptides were recognized by HCV infected individuals.

25

Evaluation of B7 immunogenicity

One of the two B7-supertype cross-reactive peptides (1145.12, Core 169) has been evaluated for immunogenicity in HCV-infected patients. Two independent collaborators have shown that this peptide is indeed immunogenic, and is recognized by T cells from
 30 HCV-infected patients (Chang *et al.*, *J. Immunol.* 162:1156-1164, 1999)

Example 4: Implementation of the Extended Supermotif to Improve the Binding Capacity of Native Epitopes by Creating Analogs

HLA motifs and supermotifs (comprising primary and/or secondary residues) are useful in the identification and preparation of highly cross-reactive native peptides, as demonstrated herein. Moreover, the definition of HLA motifs and supermotifs also allows one to engineer highly cross-reactive epitopes by identifying residues within a native peptide sequence which can be analogued, or "fixed" to confer upon the peptide certain characteristics, *e.g.* greater cross-reactivity within the group of HLA molecules that comprise a supertype, and/or greater binding affinity for some or all of those HLA molecules. Examples of analog peptides that exhibit modulated binding affinity are set forth in this example.

Analoging at Primary Anchor Residues

As shown in Example 2, more than ten different HCV-derived, A2-supertype-restricted epitopes were identified. Peptide engineering strategies are implemented to further increase the cross-reactivity of the candidate epitopes identified above which bind 3/5 of the A2 supertype alleles tested. On the basis of the data disclosed, *e.g.*, in related and co-pending U.S.S.N 09/226,775, the main anchors of A2-supermotif-bearing peptides are altered, for example, to introduce a preferred L, I, V, or M at position 2, and I or V at the C-terminus.

To analyze the cross-reactivity of the analog peptides, each engineered analog is initially tested for binding to the prototype A2 supertype allele A*0201, then, if A*0201 binding capacity is maintained, for A2-supertype cross-reactivity.

Similarly, analogs of HLA-A3 supermotif-bearing epitopes may also be generated. For example, peptides binding to 3/5 of the A3-supertype molecules may be engineered at primary anchor residues to possess a preferred residue (V, S, M, or A) at position 2.

The analog peptides are then tested for the ability to bind A*03 and A*11 (prototype A3 supertype alleles). Those peptides that demonstrate ≤ 500 nM binding capacity are then tested for A3-supertype cross-reactivity.

Similarly to the A2- and A3- motif bearing peptides, peptides binding 3 or more B7-supertype alleles may be improved, where possible, to achieve increased cross-reactive binding. B7 supermotif-bearing peptides may, for example, be engineered to

possess a preferred residue (V, I, L, or F) at the C-terminal primary anchor position, as demonstrated by Sidney *et al.* (*J. Immunol.* 157:3480-3490, 1996).

Analoging at Secondary Anchor Residues

Moreover, HLA supermotifs are of value in engineering highly cross-reactive peptides and/or peptides that bind HLA molecules with increased affinity by identifying particular residues at secondary anchor positions that are associated with such properties. Demonstrating this, the binding capacity of a peptide representing a discreet single amino acid substitution at position one was analyzed. Peptide 1145.13 (Table XXVIIIc), which represents the substitution of L to F at position 1 of the core 169 sequence, binds all five B7-supertype molecules with a good affinity (all IC_{50} values ≤ 132 nM), and in 3 instances has higher affinity over that of the parent peptide by >35-fold.

Because so few B7-supertype cross-reactive epitopes were identified, our results from previous binding evaluations were analyzed to identify conserved (8-, 9-, 10-, or 11-mer) peptides which bind, minimally, 3/5 B7 supertype molecules with weak affinity (IC_{50} of 500nM-5 μ M). This analysis identified 9 peptides, 6 of which are analogued (including core 169 which had been previously analogued). These peptides are tested for enhanced binding affinity and B7-supertype cross-reactivity.

Engineered analogs with sufficiently improved binding capacity or cross-reactivity are tested for immunogenicity in HLA-B7-transgenic mice, following for example, IFA immunization or lipopeptide immunization.

In conclusion, these data demonstrate that by the use of even single amino acid substitutions, it is possible to increase the binding affinity and/or cross-reactivity of peptide ligands for HLA supertype molecules.

Example 5: Identification of conserved HCV-derived sequences with HLA-DR binding motifs

Peptide epitopes bearing an HLA class II supermotif or motif may also be identified as outlined below using methodology similar to that described in Examples 1-3.

Selection of HLA-DR-supermotif-bearing epitopes

To identify HCV-derived, HLA class II HTL epitopes, the same fourteen HCV polyprotein sequences used for the identification of HLA Class I supermotif/motif

sequences were analyzed for the presence of sequences bearing an HLA-DR-motif or supermotif. Specifically, 15-mer sequences were selected comprising a DR-supermotif, further comprising a 9-mer core, and three-residue N- and C-terminal flanking regions (15 amino acids total). It was also required that the 15-mer sequence be conserved in at least 79% (11/14) of the HCV strains analyzed. These criteria identified a total of 49 non-redundant sequences, which are shown in Table XXXIIA. (In the context of Class II epitopes, a sequence is considered operationally redundant if more than 80% of it's sequence overlaps with another peptide.)

Protocols for predicting peptide binding to DR molecules have been developed (Southwood *et al.*, *J. Immunol.* 160:3363-3373, 1998). These protocols, specific for individual DR molecules, allow the scoring, and ranking, of 9-mer core regions. Each protocol not only scores peptide sequences for the presence of DR-supermotif primary anchors (i.e., at position 1 and position 6) within a 9-mer core, but additionally evaluates sequences for the presence of secondary anchors. Using allele specific selection tables (see, *e.g.*, Southwood *et al.*, *ibid.*), it has been found that these protocols efficiently select peptide sequences with a high probability of binding a particular DR molecule. Additionally, it has been found that performing these protocols in tandem, specifically those for DR1, DR4w4, and DR7, can efficiently select DR cross-reactive peptides.

To see if these protocols serve to identify additional epitopes, the same HCV polyproteins used above were re-scanned for the presence of 15-mer peptides with 9-mer core regions that were $\geq 79\%$ (11/14 strains) conserved. This identified 152 sequences; 49 of which were identified previously, as described above. Next, the 9-mer core region of each of these peptides was scored using the DR1, DR4w4, and DR7 algorithms. Twenty-two peptides, including 12 new sequences (10 peptides were from the original set of 49) were found to have 9-mer cores with protocol-derived scores predictive of cross-reactive DR binders. The 12 additional sequences are shown in Table XXXIIB.

The conserved, HCV-derived peptides identified above were tested for their binding capacity for various common HLA-DR molecules. All peptides were initially tested for binding to the DR molecules in the primary panel: DR1, DR4w4, and DR7. Peptides binding at least 2 of these 3 DR molecules were then tested for binding to DR2w2 β 1, DR2w2 β 2, DR6w19, and DR9 molecules in secondary assays. Finally, peptides binding at least 2 of the 4 secondary panel DR molecules, and thus cumulatively at least 4 of 7 different DR molecules, were screened for binding to DR4w15, DR5w11,

and DR8w2 molecules in tertiary assays. Peptides binding at least 7 of the 10 DR molecules comprising the primary, secondary, and tertiary screening assays were considered cross-reactive DR binders. The composition of these screening panels, and the phenotypic frequency of associated antigens, are shown in Table XXXIII.

5 Upon testing, it was found that 29 of the original 75 peptides (39%) bound two or more of the primary HLA molecules. Twenty-six of these cross-reactive binders were then tested in the secondary assays, and nineteen were found to bind at least four of the seven HLA DR molecules in the primary and secondary panels. Finally, the nineteen peptides passing the secondary screening phase were tested for binding in the tertiary
10 assays. As a result, nine peptides were identified which bound at least seven of ten common HLA-DR molecules. Table XXXIV shows these nine peptides and their binding capacity for each allele-specific HLA-DR molecule in the primary through tertiary panels. Also shown in Table XXXIV are two peptides (F134.05 and F134.08) for which a complete binding analysis was not performed. However, both of these peptides bound six
15 of the seven HLA DR molecules tested. F134.08 nests peptide 1283.44, which bound eight of 10 allele-specific HLA molecules.

 In conclusion, eleven cross-reactive DR-binding peptides, derived from six discrete (*i.e.* non-redundant) regions of the HCV genome, have been identified. Two of the six regions from which these epitopes were derived are covered by multiple,
20 overlapping epitopes.

Selection of conserved DR3 motif peptides

 Because HLA-DR3 is an allele that is prevalent in Caucasian, Black, and Hispanic populations, DR3 binding capacity is an important criterion in the selection of HTL
25 epitopes. However, data generated previously indicated that DR3 only rarely cross-reacts with other DR alleles (Sidney *et al.*, *J. Immunol.* 149:2634-2640, 1992; Geluk *et al.*, *J. Immunol.* 152:5742-5748, 1994; Southwood *et al.*, *J. Immunol.* 160:3363-3373, 1998). This is not entirely surprising in that the DR3 peptide-binding motif appears to be distinct from the specificity of most other DR alleles.

30 To efficiently identify peptides that bind DR3, target proteins were analyzed for conserved sequences carrying one of the two DR3 specific binding motifs reported by Geluk *et al.* (*J. Immunol.* 152:5742-5748, 1994). Fifteen sequences, including a peptide nested within a DR-supermotif sequence identified above (peptide Pape 22), were

identified (Table XXXIId). Preferably, DR3 motifs will be found clustered in proximity with DR supermotif regions.

Fourteen of the fifteen peptides containing a DR3 motif were tested for their DR3 binding capacity. Two peptides (CH35.0106 and CH35.0107) were found to bind DR3 with an affinity of 1 μ M or less (Table XXXV), and thereby qualify as HLA class II high affinity binders.

DR3 binding epitopes identified in this manner may then be included in vaccine compositions with DR supermotif-bearing peptide epitopes.

10 Example 6: Immunogenicity of candidate HCV-derived HTL epitopes and known dominant HCV HTL epitope

In the course of collaborative studies with G. Pape and C. Ferrari, eight conserved, HCV-derived peptides have been identified which are recognized by HCV-infected individuals.

15 One of these studies (Diepolder *et al.*, *J. Virol.* 71:6011-6019, 1997), identified peptide F98.05, which spans residues 1248-1261 of the NS3 protein, as an immunodominant CD4⁺ T-cell epitope that was recognized by 14/23 NS3-specific CD4⁺ T-cell clones from 4/5 patients with acute hepatitis C infection. This epitope, shown above to be an HLA-DR cross-reactive binder (see Table XXXIV), was capable of being
20 presented to helper CD4⁺ T cells by multiple HLA molecules (DR4, DR11, DR12, DR13, and DR16). Two other peptides, Pape 22 and Pape 29, were also recognized by CD4⁺ T cell clones, although, in a more limited context; correspondingly, neither of these peptides are DR-cross-reactive binders.

By direct peripheral blood T cell stimulation and by fine specificity analysis of
25 HCV-specific T-cell lines and clones, studies done in collaboration with Ferrari's group identified 6 immunodominant epitopes, including one also identified in the Pape collaboration, that are derived from conserved regions of the core, NS3, and NS4 proteins. These epitopes were also found to be cross-reactive, being presented to T cells in the context of different Class II molecules. Three of the 6 epitopes, F98.04 (F134.03),
30 F134.05 and F134.08, are cross-reactive HLA-DR binders (see Table XXXIV).

In conclusion, the immunogenicity of 8 epitopes derived from conserved regions of the HCV genome has been demonstrated. Three of these epitopes (F98.05, F134.05, and F134.08; see Table XXXIV) are broadly cross-reactive HLA-DR binding peptides.

Example 7. Calculation of phenotypic frequencies of HLA-supertypes in various ethnic backgrounds to determine breadth of population coverage

This example illustrates the assessment of the breadth of population coverage of a vaccine composition comprised of multiple epitopes comprising multiple supermotifs and/or motifs.

In order to analyze population coverage, gene frequencies of HLA alleles were determined. Gene frequencies for each HLA allele were calculated from antigen or allele frequencies utilizing the binomial distribution formulae $gf=1-(SQRT(1-af))$ (see, *e.g.*, Sidney *et al.*, *Human Immunol.* 45:79-93, 1996). To obtain overall phenotypic frequencies, cumulative gene frequencies were calculated, and the cumulative antigen frequencies derived by the use of the inverse formula $[af=1-(1-Cgf)^2]$.

Where frequency data was not available at the level of DNA typing, correspondence to the serologically defined antigen frequencies was assumed. To obtain total potential supertype population coverage no linkage disequilibrium was assumed, and only alleles confirmed to belong to each of the supertypes were included (minimal estimates). Estimates of total potential coverage achieved by inter-loci combinations were made by adding to the A coverage the proportion of the non-A covered population that could be expected to be covered by the B alleles considered (*e.g.*, $total=A+B*(1-A)$). Confirmed members of the A3-like supertype are A3, A11, A31, A*3301, and A*6801. Although the A3-like supertype may also include A34, A66, and A*7401, these alleles were not included in overall frequency calculations. Likewise, confirmed members of the A2-like supertype family are A*0201, A*0202, A*0203, A*0204, A*0205, A*0206, A*0207, A*6802, and A*6901. Finally, the B7-like supertype-confirmed alleles are: B7, B*3501-03, B51, B*5301, B*5401, B*5501-2, B*5601, B*6701, and B*7801 (potentially also B*1401, B*3504-06, B*4201, and B*5602).

Population coverage achieved by combining the A2-, A3- and B7-supertypes is approximately 86% in five major ethnic groups (see Table XXI). Coverage may be extended by including peptides bearing the A1 and A24 motifs. On average, A1 is present in 12% and A24 in 29% of the population across five different major ethnic groups (Caucasian, North American Black, Chinese, Japanese, and Hispanic). Together, these alleles are represented with an average frequency of 39% in these same ethnic populations. The total coverage across the major ethnicities when A1 and A24 are combined with the coverage of the A2-, A3- and B7-supertype alleles is >95%. An

analogous approach can be used to estimate population coverage achieved with combinations of class II motif-bearing epitopes.

Summary of candidate HLA class I and class II epitopes

5 In summary, on the basis of the data presented in the above examples, 26 CTL candidate peptide epitopes derived from conserved regions of the HCV virus have been identified (Table XXXVIa). These include twelve HLA-A2 supermotif-bearing epitopes, eight HLA-A3 supermotif-bearing epitopes, and one HLA-B7 supermotif-bearing epitope, each capable of binding to multiple A2-, A3-, or B7-supertype molecules, and
10 immunogenic in HLA transgenic mice or antigenic for human PBL (with the exception of peptide 29.0035/1260.04). Additional epitopes not evaluated for immunogenicity are also included. They are an additional B7-supermotif-bearing epitope and two HLA-A1 and one HLA-A24 high-affinity binding peptides. A known HLA-A31 restricted epitope (VGIYLLPNR), which also binds HLA-A33, is also set out in Table XXXVIa and is
15 useful in combination with other Class I or Class II epitopes.

 With these 26 CTL epitopes (as disclosed herein and from the art), average population coverage, (*i.e.*, recognition of at least one HCV epitope), is predicted to be greater than 95% in each of five major ethnic populations. The potential redundancy of coverage afforded by 25 of these epitopes (the peptide 24.0086 was not included) was
20 estimated using the game theory Monte Carlo simulation analysis, which is known in the art (see *e.g.*, Osborne, M.J. and Rubinstein, A. "A course in game theory" MIT Press, 1994). As shown in Figure 1, it is estimated that 90% of the individuals in a population comprised of the Caucasian, North American Black, Japanese, Chinese, and Hispanic ethnic groups would recognize 2 or more of the candidate epitopes described herein.

25 A list of HCV-derived HTL epitopes that would be preferred for use in the design of minigene constructs or other vaccine formulations is summarized in Table XXXVIb. As shown, 9 different peptide-binding regions have been identified which bind multiple HLA-DR molecules or bind HLA-DR3. (In the case of the NS4 1914-1935 region, the longer peptide, F134.08, recognized by patients, was chosen over the shorter peptide,
30 1283.44. The longer peptide essentially incorporates the shorter peptide, and also binds additional DR molecules that the shorter peptide does not bind.) Three of these peptides have been recognized as dominant epitopes in HCV infected patients.

 It is estimated that each of 10 common DR molecules recognizing the DR supermotif, and DR3, are covered by a minimum of 2 epitopes. Correspondingly, the

total estimated population coverage represented by this panel of epitopes is in excess of 91% in each of the 5 major ethnic populations (Table XXXVII).

Example 8: Recognition Of Generation Of Endogenous Processed Antigens After Priming

This example determines that CTL induced by native or analogued peptide epitopes identified and selected as described in Examples 1-6 recognize endogenously synthesized, *i.e.*, native antigens.

Effector cells isolated from transgenic mice that are immunized with peptide epitopes as in Example 3, for example HLA-A2 supermotif-bearing epitopes, are re-stimulated *in vitro* using peptide-coated stimulator cells. Six days later, effector cells are assayed for cytotoxicity and the cell lines that contain peptide-specific cytotoxic activity are further re-stimulated. An additional six days later, these cell lines are tested for cytotoxic activity on ^{51}Cr labeled Jurkat-A2.1/K^b target cells in the absence or presence of peptide, and also tested on ^{51}Cr labeled target cells bearing the endogenously synthesized antigen, *i.e.* cells that are stably transfected with HCV expression vectors.

The result will demonstrate that CTL lines obtained from animals primed with peptide epitope recognize endogenously synthesized HCV antigen. The choice of transgenic mouse model to be used for such an analysis depends upon the epitope(s) that is being evaluated. In addition to HLA-A*0201/K^b transgenic mice, several other transgenic mouse models including mice with human A11, which may also be used to evaluate A3 epitopes, and B7 alleles have been characterized and others (*e.g.*, transgenic mice for HLA-A1 and A24) are being developed. HLA-DR1 and HLA-DR3 mouse models have also been developed, which may be used to evaluate HTL epitopes.

Example 9: Activity Of CTL-HTL Conjugated Epitopes In Transgenic Mice

This example illustrates the induction of CTLs and HTLs in transgenic mice by use of an HCV CTL/HTL peptide conjugate whereby the vaccine composition comprises peptides administered to an HCV-infected patient or an individual at risk for HCV. The peptide composition can comprise multiple CTL and/or HTL epitopes. This analysis demonstrates enhanced immunogenicity that can be achieved by inclusion of one or more HTL epitopes in a vaccine composition. Such a peptide composition can comprise a lipidated HTL epitope conjugated to a preferred CTL epitope containing, for example, at

least one CTL epitope selected from Table XXVI-XXIX, or an analog of that epitope. The HTL epitope is, for example, selected from Table XXXII.

5 Lipopeptide preparation: Lipopeptides are prepared by coupling the appropriate fatty acid to the amino terminus of the resin bound peptide. A typical procedure is as follows: A dichloromethane solution of a four-fold excess of a pre-formed symmetrical anhydride of the appropriate fatty acid is added to the resin and the mixture is allowed to react for two hours. The resin is washed with dichloromethane and dried. The resin is then treated with trifluoroacetic acid in the presence of appropriate scavengers [e.g. 5% (v/v) water] for 60 minutes at 20°C. After evaporation of excess trifluoroacetic acid, the
10 crude peptide is washed with diethyl ether, dissolved in methanol and precipitated by the addition of water. The peptide is collected by filtration and dried.

Immunization procedures: Immunization of transgenic mice is performed as described (Alexander *et al.*, *J. Immunol.* 159:4753-4761, 1997). For example, A2/K^b mice, which are transgenic for the human HLA A2.1 allele and are useful for the
15 assessment of the immunogenicity of HLA-A*0201 motif- or HLA-A2 supermotif-bearing epitopes, are primed subcutaneously (base of the tail) with 0.1 ml of peptide conjugate formulated in saline, or DMSO/saline. Seven days after priming, splenocytes obtained from these animals are restimulated with syngenic irradiated LPS-activated lymphoblasts coated with peptide.

20 Cell lines: Target cells for peptide-specific cytotoxicity assays are Jurkat cells transfected with the HLA-A2.1/K^b chimeric gene (e.g., Vitiello *et al.*, *J. Exp. Med.* 173:1007, 1991)

In vitro CTL activation: One week after priming, spleen cells (30x10⁶ cells/flask) are co-cultured at 37°C with syngeneic, irradiated (3000 rads), peptide coated
25 lymphoblasts (10x10⁶ cells/flask) in 10 ml of culture medium/T25 flask. After six days, effector cells are harvested and assayed for cytotoxic activity.

Assay for cytotoxic activity: Target cells (1.0 to 1.5x10⁶) are incubated at 37°C in the presence of 200 µl of ⁵¹Cr. After 60 minutes, cells are washed three times and resuspended in R10 medium. Peptide is added where required at a concentration of 1
30 µg/ml. For the assay, 10⁴ ⁵¹Cr-labeled target cells are added to different concentrations of effector cells (final volume of 200 µl) in U-bottom 96-well plates. After a 6 hour incubation period at 37°C, a 0.1 ml aliquot of supernatant is removed from each well and radioactivity is determined in a Micromedic automatic gamma counter. The percent specific lysis is determined by the formula: percent specific release = 100 x

(experimental release - spontaneous release)/(maximum release - spontaneous release).

To facilitate comparison between separate CTL assays run under the same conditions, % ^{51}Cr release data is expressed as lytic units/ 10^6 cells. One lytic unit is arbitrarily defined as the number of effector cells required to achieve 30% lysis of 10,000 target cells in a 6 hour ^{51}Cr release assay. To obtain specific lytic units/ 10^6 , the lytic units/ 10^6 obtained in the absence of peptide is subtracted from the lytic units/ 10^6 obtained in the presence of peptide. For example, if 30% ^{51}Cr release is obtained at the effector (E): target (T) ratio of 50:1 (i.e., 5×10^5 effector cells for 10,000 targets) in the absence of peptide and 5:1 (i.e., 5×10^4 effector cells for 10,000 targets) in the presence of peptide, the specific lytic units would be: $[(1/50,000) - (1/500,000)] \times 10^6 = 18 \text{ LU}$.

The results are analyzed to assess the magnitude of the CTL responses of animals injected with the immunogenic CTL/HTL conjugate vaccine preparation and are compared to the magnitude of the CTL response achieved using the CTL epitope as outlined in Example 3. Analyses similar to this may be performed to evaluate the immunogenicity of peptide conjugates containing multiple CTL epitopes and/or multiple HTL epitopes. In accordance with these procedures it is found that a CTL response is induced, and concomitantly that an HTL response is induced upon administration of such compositions.

20 Example 10. Selection of CTL and HTL epitopes for inclusion in an HCV-specific vaccine.

This example illustrates the procedure for the selection of peptide epitopes for vaccine compositions of the invention. The peptides in the composition may be in the form of a nucleic acid sequence, either single or one or more sequences (i.e., minigene) that encodes peptide(s), or may be single and/or polyepitopic peptides.

The following principles are utilized when selecting an array of epitopes for inclusion in a vaccine composition. Each of the following principles are balanced in order to make the selection.

1.) Epitopes are selected which, upon administration, mimic immune responses that have been observed to be correlated with HCV clearance. For HLA Class I this includes 3-4 epitopes that come from at least one antigen of HCV. In other words, it has been observed that patients who spontaneously clear HCV generate an immune response

to at least 3 epitopes on at least one HCV antigen. For HLA Class II a similar rationale is employed; again 3-4 epitopes are selected from at least one HCV antigen.

2.) Epitopes are selected that have the requisite binding affinity established to be correlated with immunogenicity: for HLA Class I an IC_{50} of 500 nM or less, or for
5 Class II an IC_{50} of 1000 nM or less.

3.) Sufficient supermotif bearing peptides, or a sufficient array of allele-specific motif bearing peptides, are selected to give broad population coverage. For example, epitopes are selected to provide at least 80% population coverage. A Monte Carlo analysis, a statistical evaluation known in the art and discussed herein, can be
10 employed to assess breadth, or redundancy, of population coverage.

4.) When selecting epitopes for HCV antigens it may be preferable to select native epitopes. Therefore, of particular relevance for infectious disease vaccines, are epitopes referred to as "nested epitopes." Nested epitopes occur where at least two epitopes overlap in a given peptide sequence. A peptide comprising "transcendent nested
15 epitopes" is a peptide that has both HLA class I and HLA class II epitopes in it.

When providing nested epitopes, a sequence that has the greatest number of epitopes per provided sequence is provided. A limitation on this principle is to avoid providing a peptide that is any longer than the amino terminus of the amino terminal epitope and the carboxyl terminus of the carboxyl terminal epitope in the peptide. When
20 providing a longer peptide sequence, such as a sequence comprising nested epitopes, the sequence is screened in order to insure that it does not have pathological or other deleterious biological properties.

5.) When creating a minigene, as disclosed in greater detail in Example 11, an objective is to generate the smallest peptide possible that encompasses the epitopes of
25 interest. The principles employed are similar, if not the same as those employed when selecting a peptide comprising nested epitopes. Additionally, however, upon determination of the nucleic acid sequence to be provided as a minigene, the peptide encoded thereby is analyzed to determine whether any "junctional epitopes" have been created. A junctional epitope is an actual binding epitope, as predicted, *e.g.*, by motif
30 analysis. Junctional epitopes are generally to be avoided because the recipient may generate an immune response to that epitope, which is not present in a native HCV protein sequence. Of particular concern is a junctional epitope that is a "dominant epitope." A dominant epitope may lead to such a zealous response that immune responses to other epitopes are diminished or suppressed.

Peptide epitopes for inclusion in vaccine compositions are, for example, selected from those listed in Tables XXVI-XXIX and Table XXXII. A vaccine composition comprised of selected peptides, when administered, is safe, efficacious, and elicits an immune response similar in magnitude of an immune response that clears an acute HCV infection.

Example 11: Construction of Minigene Multi-Epitope DNA Plasmids

This example provides general guidance for the construction of a minigene expression plasmid. Minigene plasmids may, of course, contain various configurations of CTL and/or HTL epitopes or epitope analogs as described herein. Expression plasmids have been constructed and evaluated as described, for example, in co-pending U.S.S.N. 09/311,784 filed 5/13/99. An example of such a plasmid for the expression of HCV epitopes is shown in Figure 2, which illustrates the orientation of HCV peptide epitopes in a minigene construct.

A minigene expression plasmid may include multiple CTL and HTL peptide epitopes. In the present example, HLA-A2, -A3, -B7 supermotif-bearing peptide epitopes and HLA-A1 and -A24 motif-bearing peptide epitopes are used in conjunction with DR supermotif-bearing epitopes and/or DR3 epitopes (Figure 2). Preferred epitopes are identified, for example, in Tables XXVI-XXIX and XXXII. HLA class I supermotif or motif-bearing peptide epitopes derived from multiple HCV antigens, *e.g.*, the core, NS4, NS3, NS5, NS1/E2, are selected such that multiple supermotifs/motifs are represented to ensure broad population coverage. Similarly, HLA class II epitopes are selected from multiple HCV antigens to provide broad population coverage, *i.e.* both HLA DR-1-4-7 supermotif-bearing epitopes and HLA DR-3 motif-bearing epitopes are selected for inclusion in the minigene construct. The selected CTL and HTL epitopes are then incorporated into a minigene for expression in an expression vector.

This example illustrates the methods to be used for construction of such a minigene-bearing expression plasmid. Other expression vectors that may be used for minigene compositions are available and known to those of skill in the art.

The minigene DNA plasmid contains a consensus Kozak sequence and a consensus murine kappa Ig-light chain signal sequence followed by CTL and/or HTL epitopes selected in accordance with principles disclosed herein. The sequence encodes an open reading frame fused to the Myc and His antibody epitope tag coded for by the pcDNA 3.1 Myc-His vector.

Overlapping oligonucleotides, for example eight oligonucleotides, averaging approximately 70 nucleotides in length with 15 nucleotide overlaps, are synthesized and HPLC-purified. The oligonucleotides encode the selected peptide epitopes as well as appropriate linker nucleotides, Kozak sequence, and signal sequence. The final
 5 multiepitope minigene is assembled by extending the overlapping oligonucleotides in three sets of reactions using PCR. A Perkin/Elmer 9600 PCR machine is used and a total of 30 cycles are performed using the following conditions: 95°C for 15 sec, annealing temperature (5° below the lowest calculated T_m of each primer pair) for 30 sec, and 72°C for 1 min.

10 For the first PCR reaction, 5 µg of each of two oligonucleotides are annealed and extended: Oligonucleotides 1+2, 3+4, 5+6, and 7+8 are combined in 100 µl reactions containing *Pfu* polymerase buffer (1x= 10 mM KCL, 10 mM (NH₄)₂SO₄, 20 mM Tris-chloride, pH 8.75, 2 mM MgSO₄, 0.1% Triton X-100, 100 µg/ml BSA), 0.25 mM each dNTP, and 2.5 U of *Pfu* polymerase. The full-length dimer products are gel-purified, and
 15 two reactions containing the product of 1+2 and 3+4, and the product of 5+6 and 7+8 are mixed, annealed, and extended for 10 cycles. Half of the two reactions are then mixed, and 5 cycles of annealing and extension carried out before flanking primers are added to amplify the full length product for 25 additional cycles. The full-length product is gel-purified and cloned into pCR-blunt (Invitrogen) and individual clones are screened by
 20 sequencing.

Example 12. The plasmid construct and the degree to which it induces immunogenicity.

The degree to which the plasmid construct prepared using the methodology outlined in Example 11 is able to induce immunogenicity is evaluated through *in vivo*
 25 injections into mice and subsequent *in vitro* assessment of CTL and HTL activity, which are analysed using cytotoxicity and proliferation assays, respectively, as detailed *e.g.*, in U.S.S.N. 09/311,784 filed 5/13/99 and Alexander *et al.*, *Immunity* 1:751-761, 1994. To assess the capacity of the pMin minigene construct to induce CTLs *in vivo*, HLA-A11/K^b transgenic mice, for example, are immunized intramuscularly with 100 µg of naked
 30 cDNA. As a means of comparing the level of CTLs induced by cDNA immunization, a control group of animals is also immunized with an actual peptide composition that comprises multiple epitopes synthesized as a single polypeptide as they would be encoded by the minigene.

Splenocytes from immunized animals are stimulated twice with each of the respective compositions (peptide epitopes encoded in the minigene or the polyepitopic peptide), then assayed for peptide-specific cytotoxic activity in a ^{51}Cr release assay. The results indicate the magnitude of the CTL response directed against the A3-restricted epitope, thus indicating the *in vivo* immunogenicity of the minigene vaccine and polyepitopic vaccine. It is, therefore, found that the minigene elicits immune responses directed toward the HLA-A3 supermotif peptide epitopes as does the polyepitopic peptide vaccine. A similar analysis is also performed using other HLA-A2 and HLA-B7 transgenic mouse models to assess CTL induction by HLA-A2 and HLA-B7 motif or supermotif epitopes.

To assess the capacity of a class II epitope encoding minigene to induce HTLs *in vivo*, I-A^b restricted mice, for example, are immunized intramuscularly with 100 μg of plasmid DNA. As a means of comparing the level of HTLs induced by DNA immunization, a group of control animals is also immunized with an actual peptide composition emulsified in complete Freund's adjuvant.

CD4⁺ T cells, *i.e.* HTLs, are purified from splenocytes of immunized animals and stimulated with each of the respective compositions (peptides encoded in the minigene). The HTL response is measured using a ^3H -thymidine incorporation proliferation assay, (*see, e.g.*, Alexander et al. Immunity 1:751-761, 1994). the results indicate the magnitude of the HTL response, thus demonstrating the *in vivo* immunogenicity of the minigene.

Example 13: Peptide Composition for Prophylactic Uses

Vaccine compositions of the present invention are used to prevent HCV infection in persons who are at risk for such infection. For example, a polyepitopic peptide epitope composition (or a nucleic acid comprising the same) containing multiple CTL and HTL epitopes such as those selected in Examples 9 and/or 10, which are also selected to target greater than 80% of the population, is administered to individuals at risk for HCV infection. The composition is provided as a single lipidated polypeptide that encompasses multiple epitopes. The vaccine is administered in an aqueous carrier comprised of Freund's Incomplete Adjuvant. The dose of peptide for the initial immunization is from about 1 to about 50,000 μg , generally 100-5,000 μg , for a 70 kg patient. The initial administration of vaccine is followed by booster dosages at 4 weeks followed by evaluation of the magnitude of the immune response in the patient, by

techniques that determine the presence of epitope-specific CTL populations in a PBMC sample. Additional booster doses are administered as required. The composition is found to be both safe and efficacious as a prophylaxis against HCV infection.

Alternatively, the polyepitopic peptide composition can be administered as a
5 nucleic acid in accordance with methodologies known in the art and disclosed herein.

Example 14: Polyepitopic Vaccine Compositions Derived from Native HCV Sequences

A native HCV polyprotein sequence is screened, preferably using computer algorithms defined for each class I and/or class II supermotif or motif, to identify
10 "relatively short" regions of the polyprotein that comprise multiple epitopes and is preferably less in length than an entire native antigen. This relatively short sequence that contains multiple distinct, even overlapping, epitopes is selected and used to generate a minigene construct. The construct is engineered to express the peptide, which corresponds to the native protein sequence. The "relatively short" peptide is generally
15 less than 250 amino acids in length, often less than 100 amino acids in length, preferably less than 75 amino acids in length, and more preferably less than 50 amino acids in length. The protein sequence of the vaccine composition is selected because it has maximal number of epitopes contained within the sequence, *i.e.*, it has a high concentration of epitopes. As noted herein, epitope motifs may be nested or overlapping
20 (*i.e.*, frame shifted relative to one another). For example, with frame shifted overlapping epitopes, two 9-mer epitopes and one 10-mer epitope can be present in a 10 amino acid peptide. Such a vaccine composition is administered for therapeutic or prophylactic purposes.

The vaccine composition will preferably include, for example, three CTL epitopes
25 and at least one HTL epitope from HCV. This polyepitopic native sequence is administered either as a peptide or as a nucleic acid sequence which encodes the peptide. Alternatively, an analog can be made of this native sequence, whereby one or more of the epitopes comprise substitutions that alter the cross-reactivity and/or binding affinity properties of the polyepitopic peptide.

30 The embodiment of this example provides for the possibility that an as yet undiscovered aspect of immune system processing will apply to the native nested sequence and thereby facilitate the production of therapeutic or prophylactic immune response-inducing vaccine compositions. Additionally such an embodiment provides for the possibility of motif-bearing epitopes for an HLA makeup that is presently unknown.

Furthermore, this embodiment (absent analogs) directs the immune response to multiple peptide sequences that are actually present in native HCV antigens thus avoiding the need to evaluate any junctional epitopes. Lastly, the embodiment provides an economy of scale when producing nucleic acid vaccine compositions.

- 5 Related to this embodiment, computer programs can be derived in accordance with principles in the art, which identify in a target sequence, the greatest number of epitopes per sequence length.

Example 15. Polyepitopic Vaccine Compositions Directed To Multiple Diseases

- 10 The HCV peptide epitopes of the present invention are used in conjunction with peptide epitopes from target antigens related to one or more other diseases, to create a vaccine composition that is useful for the prevention or treatment of HCV as well as the one or more other disease(s). Examples of the other diseases include, but are not limited to, HIV, HBV, and HPV.

- 15 For example, a polyepitopic peptide composition comprising multiple CTL and HTL epitopes that target greater than 98% of the population may be created for administration to individuals at risk for both HCV and HIV infection. The composition can be provided as a single polypeptide that incorporates the multiple epitopes from the various disease-associated sources, or can be administered as a composition comprising
20 one or more discrete epitopes.

Example 16. Use of peptides to evaluate an immune response

- Peptides of the invention may be used to analyze an immune response for the presence of specific CTL or HTL populations directed to HCV. Such an analysis may be
25 performed in a manner as that described by Ogg *et al.*, *Science* 279:2103-2106, 1998. In the following example, peptides in accordance with the invention are used as a reagent for diagnostic or prognostic purposes, not as an immunogen.

- In this example highly sensitive human leukocyte antigen tetrameric complexes ("tetramers") are used for a cross-sectional analysis of, for example, HCV HLA-A*0201-
30 specific CTL frequencies from HLA A*0201-positive individuals at different stages of infection or following immunization using an HCV peptide containing an A*0201 motif. Tetrameric complexes are synthesized as described (Musey *et al.*, *N. Engl. J. Med.* 337:1267, 1997). Briefly, purified HLA heavy chain (A*0201 in this example) and β 2-

microglobulin are synthesized by means of a prokaryotic expression system. The heavy chain is modified by deletion of the transmembrane-cytosolic tail and COOH-terminal addition of a sequence containing a BirA enzymatic biotinylation site. The heavy chain, β 2-microglobulin, and peptide are refolded by dilution. The 45-kD refolded product is isolated by fast protein liquid chromatography and then biotinylated by BirA in the presence of biotin (Sigma, St. Louis, Missouri), adenosine 5'triphosphate and magnesium. Streptavidin-phycoerythrin conjugate is added in a 1:4 molar ratio, and the tetrameric product is concentrated to 1 mg/ml. The resulting product is referred to as tetramer-phycoerythrin.

For the analysis of patient blood samples, approximately one million PBMCs are centrifuged at 300g for 5 minutes and resuspended in 50 μ l of cold phosphate-buffered saline. Tri-color analysis is performed with the tetramer-phycoerythrin, along with anti-CD8-Tricolor, and anti-CD38. The PBMCs are incubated with tetramer and antibodies on ice for 30 to 60 min and then washed twice before formaldehyde fixation. Gates are applied to contain >99.98% of control samples. Controls for the tetramers include both A*0201-negative individuals and A*0201-positive uninfected donors. The percentage of cells stained with the tetramer is then determined by flow cytometry. The results indicate the number of cells in the PBMC sample that contain epitope-restricted CTLs, thereby readily indicating the extent of immune response to the HCV epitope, and thus the stage of infection with HCV, the status of exposure to HCV, or exposure to a vaccine that elicits a protective or therapeutic response.

Example 17: Use of Peptide Epitopes to Evaluate Recall Responses

The peptide epitopes of the invention are used as reagents to evaluate T cell responses, such as acute or recall responses, in patients. Such an analysis may be performed on patients who have recovered from infection, who are chronically infected with HCV, or who have been vaccinated with an HCV vaccine.

For example, the class I restricted CTL response of persons who have been vaccinated may be analyzed. The vaccine may be any HCV vaccine. PBMC are collected from vaccinated individuals and HLA typed. Appropriate peptide epitopes of the invention that are preferably highly conserved and, optimally, bear supermotifs to provide cross-reactivity with multiple HLA supertype family members, are then used for analysis of samples derived from individuals who bear that HLA type.

PBMC from vaccinated individuals are separated on Ficoll-Histopaque density gradients (Sigma Chemical Co., St. Louis, MO), washed three times in HBSS (GIBCO Laboratories), resuspended in RPMI-1640 (GIBCO Laboratories) supplemented with L-glutamine (2mM), penicillin (50U/ml), streptomycin (50 µg/ml), and Hepes (10mM) containing 10% heat-inactivated human AB serum (complete RPMI) and plated using microculture formats. A synthetic peptide comprising an epitope of the invention is added at 10 µg/ml to each well and HBV core 128-140 epitope is added at 1 µg/ml to each well as a source of T cell help during the first week of stimulation.

In the microculture format, 4×10^5 PBMC are stimulated with peptide in 8 replicate cultures in 96-well round bottom plate in 100 µl/well of complete RPMI. On days 3 and 10, 100 ml of complete RPMI and 20 U/ml final concentration of rIL-2 are added to each well. On day 7 the cultures are transferred into a 96-well flat-bottom plate and restimulated with peptide, rIL-2 and 10^5 irradiated (3,000 rad) autologous feeder cells. The cultures are tested for cytotoxic activity on day 14. A positive CTL response requires two or more of the eight replicate cultures to display greater than 10% specific ^{51}Cr release, based on comparison with uninfected control subjects as previously described (Rehermann, *et al.*, *Nature Med.* 2:1104,1108, 1996; Rehermann *et al.*, *J. Clin. Invest.* 97:1655-1665, 1996; and Rehermann *et al.* *J. Clin. Invest.* 98:1432-1440, 1996).

Target cell lines are autologous and allogeneic EBV-transformed B-LCL that are either purchased from the American Society for Histocompatibility and Immunogenetics (ASHI, Boston, MA) or established from the pool of patients as described (Guilhot, *et al.* *J. Virol.* 66:2670-2678, 1992).

Cytotoxicity assays are performed in the following manner. Target cells consist of either allogeneic HLA-matched or autologous EBV-transformed B lymphoblastoid cell line that are incubated overnight with the synthetic peptide epitope of the invention at 10 µM, and labeled with 100 µCi of ^{51}Cr (Amersham Corp., Arlington Heights, IL) for 1 hour after which they are washed four times with HBSS.

Cytolytic activity is determined in a standard 4-h, split well ^{51}Cr release assay using U-bottomed 96 well plates containing 3,000 targets/well. Stimulated PBMC are tested at effector/target (E/T) ratios of 20-50:1 on day 14. Percent cytotoxicity is determined from the formula: $100 \times [(\text{experimental release} - \text{spontaneous release}) / (\text{maximum release} - \text{spontaneous release})]$. Maximum release is determined by lysis of targets by detergent (2% Triton X-100; Sigma Chemical Co., St. Louis, MO). Spontaneous release is <25% of maximum release for all experiments.

The results of such an analysis indicate the extent to which HLA-restricted CTL populations have been stimulated by previous exposure to HCV or an HCV vaccine.

The class II restricted HTL responses may also be analyzed. Purified PBMC are cultured in a 96-well flat bottom plate at a density of 1.5×10^5 cells/well and are stimulated with 10 $\mu\text{g/ml}$ synthetic peptide, whole antigen, or PHA. Cells are routinely plated in replicates of 4-6 wells for each condition. After seven days of culture, the medium is removed and replaced with fresh medium containing 10U/ml IL-2. Two days later, 1 μCi ^3H -thymidine is added to each well and incubation is continued for an additional 18 hours. Cellular DNA is then harvested on glass fiber mats and analyzed for ^3H -thymidine incorporation. Antigen-specific T cell proliferation is calculated as the ratio of ^3H -thymidine incorporation in the presence of antigen divided by the ^3H -thymidine incorporation in the absence of antigen.

Example 18: Induction Of Specific CTL Response In Humans

A human clinical trial for an immunogenic composition comprising CTL and HTL epitopes of the invention is set up as an IND Phase I, dose escalation study and carried out as a randomized, double-blind, placebo-controlled trial. Such a trial is designed, for example, as follows:

A total of about 27 subjects are enrolled and divided into 3 groups:

Group I: 3 subjects are injected with placebo and 6 subjects are injected with 5 μg of peptide composition;

Group II: 3 subjects are injected with placebo and 6 subjects are injected with 50 μg peptide composition;

Group III: 3 subjects are injected with placebo and 6 subjects are injected with 500 μg of peptide composition.

After 4 weeks following the first injection, all subjects receive a booster inoculation at the same dosage.

The endpoints measured in this study relate to the safety and tolerability of the peptide composition as well as its immunogenicity. Cellular immune responses to the peptide composition are an index of the intrinsic activity of this the peptide composition, and can therefore be viewed as a measure of biological efficacy. The following summarize the clinical and laboratory data that relate to safety and efficacy endpoints.

Safety: The incidence of adverse events is monitored in the placebo and drug treatment group and assessed in terms of degree and reversibility.

Evaluation of Vaccine Efficacy: For evaluation of vaccine efficacy, subjects are bled before and after injection. Peripheral blood mononuclear cells are isolated from
5 fresh heparinized blood by Ficoll-Hypaque density gradient centrifugation, aliquoted in freezing media and stored frozen. Samples are assayed for CTL and HTL activity.

The vaccine is found to be both safe and efficacious.

Example 19: Phase II Trials In Patients Infected With HCV

10 Phase II trials are performed to study the effect of administering the CTL-HTL peptide compositions to patients having chronic HCV infection. The main objectives of the trials are to determine an effective dose and regimen for inducing CTLs in chronically infected HCV patients, to establish the safety of inducing a CTL and HTL response in
15 these patients, and to see to what extent activation of CTLs improves the clinical picture of chronically infected CTL patients, as manifested by a transient flare in alanine aminotransferase (ALT), normalization of ALT, and reduction in HCV DNA. Such a study is designed, for example, as follows:

The studies are performed in multiple centers. The trial design is an open-label, uncontrolled, dose escalation protocol wherein the peptide composition is administered as
20 a single dose followed six weeks later by a single booster shot of the same dose. The dosages are 50, 500 and 5,000 micrograms per injection. Drug-associated adverse effects (severity and reversibility) are recorded.

There are three patient groupings. The first group is injected with 50 micrograms of the peptide composition and the second and third groups with 500 and 5,000
25 micrograms of peptide composition, respectively. The patients within each group range in age from 21-65, include both males and females, and represent diverse ethnic backgrounds. All of them are infected with HCV for over five years and are HIV, HBV and delta hepatitis virus (HDV) negative, but have positive levels of HCV antigen.

The magnitude and incidence of ALT flares and the levels of HCV DNA in the
30 blood are monitored to assess the effects of administering the peptide compositions. The levels of HCV DNA in the blood are an indirect indication of the progress of treatment. The vaccine composition is found to be both safe and efficacious in the treatment of chronic HCV infection.

Example 20: Alternative Method of Identifying Motif-Bearing Peptides

Another way of identifying motif-bearing peptides is to elute them from cells bearing defined MHC molecules. For example, EBV transformed B cell lines used for tissue typing, have been extensively characterized to determine which HLA molecules they express. In certain cases these cells express only a single type of HLA molecule. These cells can then be infected with a pathogenic organism, *e.g.*, HCV, HIV, *etc.* or transfected with nucleic acids that express the antigen of interest. Thereafter, peptides produced by endogenous antigen processing of peptides produced consequent to infection (or as a result of transfection) will bind to HLA molecules within the cell and be transported and displayed on the cell surface.

The peptides are then eluted from the HLA molecules by exposure to mild acid conditions and their amino acid sequence determined, *e.g.*, by mass spectral analysis (*e.g.*, Kubo *et al.*, *J. Immunol.* 152:3913, 1994). Because, as disclosed herein, the majority of peptides that bind a particular HLA molecule are motif-bearing, this is an alternative modality for obtaining the motif-bearing peptides correlated with the particular HLA molecule expressed on the cell.

Alternatively, cell lines that do not express any endogenous HLA molecules can be transfected with an expression construct encoding a single HLA allele. These cells may then be used as described, *i.e.*, they may be infected with a pathogenic organism or transfected with nucleic acid encoding an antigen of interest to isolate peptides corresponding to the pathogen or antigen of interest that have been presented on the cell surface. Peptides obtained from such an analysis will bear motif(s) that correspond to binding to the single HLA allele that is expressed in the cell.

As appreciated by one in the art, one can perform a similar analysis on a cell bearing more than one HLA allele and subsequently determine peptides specific for each HLA allele expressed. Moreover, one of skill would also recognize that means other than infection or transfection, such as loading with a protein antigen, can be used to provide a source of antigen to the cell.

The above examples are provided to illustrate the invention but not to limit its scope. For example, the human terminology for the Major Histocompatibility Complex, namely HLA, is used throughout this document. It is to be appreciated that these principles can be extended to other species as well. Thus, other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the

appended claims. All publications, patents, and patent application cited herein are hereby incorporated by reference for all purposes.

TABLE I

SUPERMOTIFS	POSITION	POSITION	POSITION
	2 (Primary Anchor)	3 (Primary Anchor)	C Terminus (Primary Anchor)
A1	T <i>ILVMS</i>		FWY
A2	L <i>IVMATQ</i>		I <i>VMATL</i>
A3	V <i>SMATLI</i>		RK
A24	Y <i>FWIVLMT</i>		FIY <i>WLM</i>
B7	P		V <i>ILFMWYA</i>
B27	RHK		F <i>YLWMIVA</i>
B44	E <i>D</i>		FWY <i>LIMVA</i>
B58	A <i>TS</i>		FWY <i>LIVMA</i>
B62	Q <i>LIVMP</i>		FWY <i>MIVLA</i>
MOTIFS			
A1	TSM		Y
A1		DEAS	Y
A2.1	L <i>MVQIAT</i>		V <i>LIMAT</i>
A3	L <i>MVISATFCGD</i>		KYR <i>HFA</i>
A11	V <i>TMLISAGNCDF</i>		KRYH
A24	Y <i>FWM</i>		FLIW
A*3101	M <i>VTALIS</i>		RK
A*3301	M <i>VALFIST</i>		RK
A*6801	A <i>VTMSLI</i>		RK
B*0702	P		L <i>MFWYAIW</i>
B*3501	P		L <i>MFWYIVA</i>
B51	P		L <i>IVFWYAM</i>
B*5301	P		I <i>MFWYALV</i>
B*5401	P		A <i>TIVLMFWY</i>

Bolded residues are preferred, italicized residues are less preferred: A peptide is considered motif-bearing if it has primary anchors at each primary anchor position for a motif or supermotif as specified in the above table.

POSITION

	1	2	3	4	5	6	7	8	C-terminus
SUPERMOTIFS									
A1	$\frac{1^\circ \text{Anchor}}{\text{TLLVMS}}$								$\frac{1^\circ \text{Anchor}}{\text{FWY}}$
A2	$\frac{1^\circ \text{Anchor}}{\text{LIVMATQ}}$								$\frac{1^\circ \text{Anchor}}{\text{LIVMAT}}$
A3	preferred	$\frac{1^\circ \text{Anchor}}{\text{VSMATLI}}$	YFW (4/5)	YFW (3/5)	YFW (4/5)	P (4/5)	$\frac{1^\circ \text{Anchor}}{\text{RK}}$		
	deleterious	DE (3/5); P (5/5)	DE (4/5)						
A24	$\frac{1^\circ \text{Anchor}}{\text{YFW/LM}}$ $\frac{1^\circ \text{Anchor}}{\text{YFW/LM}}$								$\frac{1^\circ \text{Anchor}}{\text{FIYWL/M}}$
B7	preferred	FWY (5/5) LIVM (3/5)	$\frac{1^\circ \text{Anchor}}{\text{P}}$	FWY (4/5)		FWY (3/5)	$\frac{1^\circ \text{Anchor}}{\text{VILFM/WYA}}$		
	deleterious	DE (3/5); P (5/5); G (4/5); A (3/5); QN (3/5)		DE (3/5)	G (4/5)	QN (4/5)	DE (4/5)		
B27	$\frac{1^\circ \text{Anchor}}{\text{RIHK}}$								$\frac{1^\circ \text{Anchor}}{\text{FYLM/IVA}}$
B44	$\frac{1^\circ \text{Anchor}}{\text{FID}}$								$\frac{1^\circ \text{Anchor}}{\text{FWYLMVA}}$
B58	$\frac{1^\circ \text{Anchor}}{\text{ATIS}}$								$\frac{1^\circ \text{Anchor}}{\text{FWYLMVA}}$
B62	$\frac{1^\circ \text{Anchor}}{\text{QI/IVA/P}}$								$\frac{1^\circ \text{Anchor}}{\text{FWYML/VA}}$

POSITION

1 2 3 4 5 6 7 8 C-terminus

POSITION

1 2 3 4 5 6 7 8 C-terminus

MOTIFS

A1 preferred GIFYW 1°Anchor DE A YFW P DEQN YFW 1°Anchor
9-mer STM

94

deleterious DE RHKLIVM A G A

A1 preferred GRIHK ASTCLIV 1°Anchor GSTC ASTC LIVM DE 1°Anchor
9-mer deleterious A RHKDEPY FW DE PQN RHK PG GP Y

POSITION

	1	2	3	4	5	6	7	8	9	C-terminus
									or	C-terminus
A1 10-mer	preferred YFW	<u>1°Anchor</u> STM	DEAQN	A	YFWQN		PASTC	GDE	P	<u>1°Anchor</u> Y
deleterious	GIP		RHKGLIV M	DE	RHK	QNA	RHKYFW	RHK	A	

A1 10-mer	preferred YFW	STCLIVM	<u>1°Anchor</u> DEAS	A	YFW		PG	G	YFW	<u>1°Anchor</u> Y
deleterious	RHK	RHKDEPY FW		P	G		PRHK	QN		

95

A2.1 9-mer	preferred YFW	<u>1°Anchor</u> LMIVQAT	YFW	STC	YFW		A	P	<u>1°Anchor</u> VLIMAT
deleterious	DEP		DERKH			RKH	DERKH		

A2.1 10-mer	preferred AYFW	<u>1°Anchor</u> LMIVQAT	LVIM	G		G		FYWL VIM	<u>1°Anchor</u> VLIMAT
deleterious	DEP		DE	RKHA	P		RKH	DERK H	RKH

POSITION

1 2 3 4 5 6 7 8 9 C-terminus
or terminus

A3 preferred RHHK $\frac{1^\circ \text{Anchor}}{\text{LMVISAT}}$ YFW PRHKYFW A YFW P $\frac{1^\circ \text{Anchor}}{\text{KYRHFA}}$

deleterious DEIP DE

A11 preferred A $\frac{1^\circ \text{Anchor}}{\text{VTLMISA GNCDF}}$ YFW YFW A YFW YFW P $\frac{1^\circ \text{Anchor}}{\text{KRYH}}$

deleterious DEIP A G

A24 preferred YFWRHHK $\frac{1^\circ \text{Anchor}}{\text{YFWM}}$ STC YFW YFW $\frac{1^\circ \text{Anchor}}{\text{FLIW}}$

deleterious DEIG DE G QNP DERHHK G AQN

A24 preferred $\frac{1^\circ \text{Anchor}}{\text{YFWM}}$ P YFWP P $\frac{1^\circ \text{Anchor}}{\text{FLIW}}$

deleterious GIDE QN RHH DE A QN DEFA

A3101 preferred RHHK $\frac{1^\circ \text{Anchor}}{\text{MVTALIS}}$ YFW P YFW YFW AP $\frac{1^\circ \text{Anchor}}{\text{RK}}$

POSITION

1
2
3
4
5
6
7
8
9
C-terminus
or
terminus

deleterious DEP

DE

ADE

DE

DE

DE

A3301 preferred

$\frac{1^\circ \text{Anchor}}{\text{MVALFIS}}$
7

YFW

AYFW

$\frac{1^\circ \text{Anchor}}{\text{RK}}$

deleterious GP

DE

97

A6801 preferred

$\frac{1^\circ \text{Anchor}}{\text{AVTMSLI}}$

YFWLIV
M

YFW

P

$\frac{1^\circ \text{Anchor}}{\text{RK}}$

deleterious GP

DEG

RHK

A

B0702 preferred

$\frac{1^\circ \text{Anchor}}{\text{P}}$

RHK

RHK

RHK

RHK

PA

$\frac{1^\circ \text{Anchor}}{\text{LMFWYAIT}}$

deleterious DEQNP

DEP

DE

DE

GDE

QN

DE

B3501 preferred

$\frac{1^\circ \text{Anchor}}{\text{P}}$

FWY

FWY

$\frac{1^\circ \text{Anchor}}{\text{LMFWYIT/A}}$

deleterious AGP

G

G

POSITION

	1	2	3	4	5	6	7	8	9	C-terminus or terminus
B51 preferred	LIVMF $\overline{\text{WY}}$	$\overline{\text{I}^{\circ}\text{Anchor}}$ P	FWY	STC	FWY	G	G	FWY	$\overline{\text{I}^{\circ}\text{Anchor}}$ LIVFHYAM	
deleterious	AGPDERHKSIC				DE	G	DEQN	GDE		

B5301 preferred	LIVMF $\overline{\text{WY}}$	$\overline{\text{I}^{\circ}\text{Anchor}}$ P	FWY	STC	FWY	LIVMF $\overline{\text{WY}}$	FWY	$\overline{\text{I}^{\circ}\text{Anchor}}$ IMFWYALV		
deleterious	AGPQN				G	RHKQN	DE			

98

B5401 preferred	FWY	$\overline{\text{I}^{\circ}\text{Anchor}}$ P	FWYLI $\overline{\text{VM}}$	LIVM	ALIVM	FWYAP	$\overline{\text{I}^{\circ}\text{Anchor}}$ ATIVLMFW Y			
deleterious	GPQNDE		GDESTC	RHKDE	DE	QNDGE	DE			

Italicized residues indicate less preferred or "tolerated" residues.
The information in Table II is specific for 9-mers unless otherwise specified.

TABLE III

MOTIFS	POSITION								
	1° anchor 1	2	3	4	5	1° anchor 6	7	8	9
DR4 preferred	FM \bar{Y} LIVW	M	T		I	VSTCPALIM	MH		MH
deleterious				W			R		WDE
DR1 preferred	MF \bar{L} IVWY			PAMQ		VMATSP \bar{L} IC	M		AVM
deleterious		C	CH	FD	CWD		GDE	D	
DR7 preferred	MF \bar{L} IVWY	M	W	A		IVMSACTPL	M		IV
deleterious		C		G			GRD	N	G
DR Supermotif	MF \bar{L} IVWY					VMSTACP \bar{L} I			

66

DR3 MOTIFS

1° anchor 1

2

3

1° anchor 4

5

1° anchor 6

motif a
preferred

LIVMFY

D

motif b
preferred

LIVMFAY

DNQEST

KRH

Italicized residues indicate less preferred or "tolerated" residues.

S1000241

Table IV. HLA Class I Standard Peptide Binding Affinity.

ALLELE	STANDARD PEPTIDE	SEQUENCE	STANDARD BINDING AFFINITY (nM)
A*0101	944.02	YLEPAIAKY	25
A*0201	941.01	FLPSDYFPSV	5.0
A*0202	941.01	FLPSDYFPSV	4.3
A*0203	941.01	FLPSDYFPSV	10
A*0205	941.01	FLPSDYFPSV	4.3
A*0206	941.01	FLPSDYFPSV	3.7
A*0207	941.01	FLPSDYFPSV	23
A*6802	1141.02	FTQAGYPAL	40
A*0301	941.12	KVFPYALINK	11
A*1101	940.06	AVDLYHFLK	6.0
A*3101	941.12	KVFPYALINK	18
A*3301	1083.02	STLPETYVVRR	29
A*6801	941.12	KVFPYALINK	8.0
A*2402	979.02	AYIDNYNKF	12
B*0702	1075.23	APRTLVLVLL	5.5
B*3501	1021.05	FPFKYAAAF	7.2
B51	1021.05	FPFKYAAAF	5.5
B*5301	1021.05	FPFKYAAAF	9.3
B*5401	1021.05	FPFKYAAAF	10

Table V. HLA Class II Standard Peptide Binding Affinity.

Allele	Nomenclature	Standard Peptide	Sequence	Binding Affinity (nM)
DRB1*0101	DR1	515.01	PKYVKQNTLKLAT	5.0
DRB1*0301	DR3	829.02	YKTIAFDEEARR	300
DRB1*0401	DR4w4	515.01	PKYVKQNTLKLAT	45
DRB1*0404	DR4w14	717.01	YARFQSQTTLKQKT	50
DRB1*0405	DR4w15	717.01	YARFQSQTTLKQKT	38
DRB1*0701	DR7	553.01	QYIKANSKFIGITE	25
DRB1*0802	DR8w2	553.01	QYIKANSKFIGITE	49
DRB1*0803	DR8w3	553.01	QYIKANSKFIGITE	1600
DRB1*0901	DR9	553.01	QYIKANSKFIGITE	75
DRB1*1101	DR5w11	553.01	QYIKANSKFIGITE	20
DRB1*1201	DR5w12	1200.05	EALIHQLKINPYVLS	298
DRB1*1302	DR6w19	650.22	QYIKANAKFIGITE	3.5
DRB1*1501	DR2w2 β 1	507.02	GRTQDENPVVHFFKNIV TPRTPPP	9.1
DRB3*0101	DR52a	511	NGQIGNDPNRDIL	470
DRB4*0101	DRw53	717.01	YARFQSQTTLKQKT	58
DRB5*0101	DR2w2 β 2	553.01	QYIKANSKFIGITE	20

The "Nomenclature" column lists the allelic designations used in Tables XIX and XX.

Table VI

HLA-supertype	Allele-specific HLA-supertype members	
	Verified ^a	Predicted ^b
A1	A*0101, A*2501, A*2601, A*2602, A*3201	A*0102, A*2604, A*3601, A*4301, A*8001
A2	A*0201, A*0202, A*0203, A*0204, A*0205, A*0206, A*0207, A*0209, A*0214, A*6302, A*6901	A*0208, A*0210, A*0211, A*0212, A*0213
A3	A*0301, A*1101, A*3101, A*3301, A*6801	A*0302, A*1102, A*2603, A*3302, A*3303, A*3401, A*3402, A*6601, A*6602, A*7401
A24	A*2301, A*2402, A*3001	A*2403, A*2404, A*3002, A*3003
B7	B*0702, B*0703, B*0704, B*0705, B*1508, B*3501, B*3502, B*3503, B*3504, B*3505, B*3506, B*3507, B*3508, B*5101, B*5102, B*5103, B*5104, B*5105, B*5301, B*5401, B*5501, B*5502, B*5601, B*5602, B*6701, B*7801	B*1511, B*4201, B*5901
B27	B*1401, B*1402, B*1509, B*2702, B*2703, B*2704, B*2705, B*2706, B*3801, B*3901, B*3902, B*7301	B*2701, B*2707, B*2708, B*3802, B*3903, B*3904, B*3905, B*4801, B*4802, B*1510, B*1518, B*1503
B44	B*1801, B*1802, B*3701, B*4402, B*4403, B*4404, B*4001, B*4002, B*4006	B*4101, B*4501, B*4701, B*4901, B*5001
B58	B*5701, B*5702, B*5801, B*5802, B*1516, B*1517	
B62	B*1501, B*1502, B*1513, B*5201	B*1301, B*1302, B*1504, B*1505, B*1506, B*1507, B*1515, B*1520, B*1521, B*1512, B*1514, B*1519

a. Verified alleles includes alleles whose specificity has been determined by pool sequencing analysis, peptide binding assays, or by analysis of the sequences of CTL epitopes.

b. Predicted alleles are alleles whose specificity is predicted on the basis of B and F pocket structure to overlap with the supertype specificity.

Table VII

HCV A01 Super Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0101
ATGNIPOCSF	165	10	13	93	
ATLGFQAY	1265	8	14	100	
AVQMMNRIIAF	1917	11	14	100	
CIOGSSQLY	1128	9	11	79	0.3700
CTRGVAKAVDF	1190	11	11	79	
CIWMNSTGF	555	9	11	79	
CVIQTVD	1462	8	12	86	
DELVISITW	1857	9	12	86	
ETMRSPVF	1207	9	12	86	
FSYDTRCF	2870	8	11	79	
FTEAMTRY	2792	8	14	100	
FTGLTHIDAF	1567	11	13	93	
GLPQDQHLEF	1552	11	12	86	
GLSATSLISY	2921	10	11	79	0.0029
GLTHIDAF	1569	9	13	93	
GSSYGRQY	2641	8	11	79	
GTFPINAY	2063	8	11	79	
GVAGALVAF	1863	9	12	86	
GVAKAVDF	1193	8	11	79	
GVLAALAAV	1670	9	12	86	
GVIVCEKMAV	2619	11	14	100	
GVHMEIGVNY	154	11	12	86	
HLKQINDVQY	696	11	11	79	
HMMNRSGLQY	1769	11	13	93	
HMSRGEAVQW	1910	11	11	79	
IMAKNEVF	2591	8	12	86	
ITYSTGKF	1236	8	12	86	
IVDVQVLY	701	8	12	86	
KSTIKVPAAY	1241	9	12	86	0.0130
KVIDLTGCF	121	10	12	86	
LIEANILW	2235	8	12	86	
LINTKGSW	414	8	11	79	
LIPITAV	1030	8	14	100	
LLPNIIGW	1812	9	12	86	
LLSPGSRPSW	97	11	11	79	
LSAFSLHSY	2922	9	11	79	0.8100
LSPRGSRPSW	98	10	11	79	
LTCGFADMGY	126	11	12	86	
LTHIDAF	1570	8	13	93	
LVDLIAGY	1853	8	11	79	
MILMTHEE	2876	8	12	86	
NIVDVQVLY	700	9	12	86	0.0980
NLPQCSFIF	168	10	13	93	
NICVQTQVDF	1480	10	12	86	
NNRRVQDVDF	14	11	11	79	

HICV A01 Super Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0101
MDQAVGV	1108	9	11	79	
PTVSTYQKF	1295	10	11	79	
PMGFSYDTRCF	2667	11	11	79	
PSVAATLGF	1281	9	14	100	
PILVGPTRILY	1621	11	11	79	
PVQDILIEF	1554	9	12	86	
PVQDILIEFW	1554	10	12	88	
QVDFSLDPTF	1465	11	12	86	
RLVGLSAF	2918	8	12	86	
RLAPITAY	1029	9	12	86	
RMAYDMMMNW	317	10	12	86	
RMILMTTF	2875	8	12	86	
RMILMTTF	2875	9	12	86	
RVCEKMALY	2621	9	14	100	
RMEDXSNV	156	9	12	86	
STKVPAAV	1242	8	12	86	
SVAAITGF	1262	8	14	100	
SVAAITLGFAY	1262	11	14	100	
TIMAKNEVF	2590	9	11	79	
ILVGPTRILY	1622	10	11	79	
ILVPMGGV	1811	10	12	88	
TTIMAKNEVF	2509	10	11	79	
ITMRSPVF	1208	8	12	86	
IVDFSLDPTF	1466	10	12	86	
VIDLTGQF	122	9	12	86	
VLAAALAY	1671	8	12	86	
VLEDXSNV	167	8	12	88	
VIVDILAGY	1852	9	11	79	
VMGSSYGF	2639	8	11	79	
VMGSSYGFQY	2639	10	11	79	
WMNRLIAF	1920	8	14	100	
YSPQGMFE	2648	9	11	79	
YTNWQDILGVN	1106	11	11	79	
YVGLQSSVF	276	10	12	86	
79		2			

0.0300

Table VIII

HCV A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
83	13	1904	AAIIRHIV					
86	12	1673	ALAAVCL					
79	11	1250	AAQYKVL					
79	11	1250	AAQYKVL					
79	11	1250	AAQYKVL					
79	11	147	AAALAHGV					
79	11	147	AAALAHGV					
100	14	1264	AATLGFGA					
93	13	1264	AATLGFGAM					
86	12	1187	AAVCTRGV					
79	11	1187	AAVCTRGV					
79	11	1187	AAVCTRGV					
93	13	1890	AILSPGAL					
86	12	1880	AILSPGALV	0.0014				
100	14	1890	AILSPGALV	0.0035				
100	14	150	ALAHGVIV					
100	14	150	ALAHGVIV	0.0037				
86	12	1737	ALGLQIA					
86	12	609	ALSTGLHL	0.0160				
79	11	1896	ALVGVWCA	0.0010				
79	11	1896	ALVGVWCA					
79	11	1896	ALVGVWCA					
79	11	1896	ALVGVWCA					
86	12	1602	AAVPPSWDCM					
79	11	1251	AAQYKVL					
79	11	1251	AAQYKVL					
86	12	77	AAQYKVL					
86	12	1285	AATLGFAYM					
93	13	1354	AATLGFAYM					
79	11	1596	AAVCTRGV					
78	11	1419	AAVCTRGV					
100	14	1419	AAVCTRGV					
100	14	1419	AAVCTRGV	0.0002				
79	11	1188	AAVCTRGV					
79	11	1188	AAVCTRGV					
79	11	1188	AAVCTRGV					
79	11	1188	AAVCTRGV					
100	14	1917	AAVCTRGV					
100	14	1917	AAVCTRGV	0.0001				
100	14	1917	AAVCTRGV					
93	13	1903	CAALRRHV					
79	11	1530	CAWYELTPA					
86	12	2941	CLPKLGVPPL	0.0002				
86	12	739	CLWMMLLI					
79	11	1653	CMSADLEY					

HCY A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
79	11	1653	CMSADLEVT	0.0067				
79	11	1653	CTGSSDI					
79	11	1128	CTGSSDIYL					
79	11	1128	CICGSSDIYL					
79	11	1190	CTHGVAKA					
79	11	1190	CTHGVAKAV					
79	11	555	CTWMNSTGFT	0.0006				
86	12	1462	CYTOTVDFSL					
79	11	1527	DAAGAWVEL					
100	14	1574	DAIFLSOT					
86	12	1855	DILAGYGA	0.0002				
79	11	1855	DILAGYGAGV					
79	11	1855	DILAGYGAGVA					
86	12	279	DICGSVIL	0.0007				
79	11	279	DICGSVILV					
86	12	1657	DLEVVTST	0.0002				
86	12	1657	DLEVVTSTW					
86	12	1657	DLEVVTSTWL					
93	13	2617	-DGMVGEKMA-					
83	13	2617	DGVNCEKMA					
79	11	132	DLMGYPL	0.0630	0.0009	0.0490	0.0077	3.3000
79	11	132	DLMGYPLV					
79	11	132	DLMGYPLVGA					
79	11	2412	DLSGWSI	0.0008				
79	11	2412	DSGWSI					
79	11	1883	DLVNLPA	0.0001				
79	11	1883	DLVNLPAI	0.0001				
79	11	1883	DLVNLPAI					
79	11	2772	DLVICSJA	0.0001				
86	12	1134	DLVITRIIA					
86	12	1134	DLVITRIADV					
86	12	321	DMMMWWSPT					
86	12	1339	DOAETAGA					
86	12	1339	DOAETAGAIL					
86	12	1339	DOAETAGAILV					
86	12	994	DTAAGDI					
86	12	994	DTAAGDII					
86	12	124	DITICGA					
86	12	124	DITICGADL					
86	12	124	DITICGADLM					
93	13	2673	DMCTDST					

HCV A92 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
93	13	2673	DTHCFDSIV					
93	13	2673	DTHCFDSIVI					
86	12	21	DWHTGGGCI	0.0001				
86	12	21	DWHTPGGGCI					
79	11	750	EALENLV					
100	14	2794	EAMIRYSA					
86	12	2237	ENLLWROEM					
93	13	1377	ELPYGKA	0.0001				
93	13	1377	ELPYGKAI	0.0002				
100	14	2814	ELTSCSNW					
79	11	666	ELSPILLST					
79	11	666	ELSPILLSTI					
86	12	2245	EMGNITIV	0.0003				
86	12	1731	EOROKAL					
86	12	1731	EOROKAAGL					
86	12	1731	EOROKAAGIL					
86	12	1342	ETAGARLV					
86	12	1342	ETAGARLVV					
86	12	1342	ETAGARLVVL					
86	12	1342	ETAGARLVVLA					
86	12	1207	ETIMRSPV					
86	12	1207	ETIMRSPV-F					
86	12	1659	EVTSTWV	0.0001				
86	12	1659	EVTSTWVL	0.0004				
86	12	1659	EVTSTWVLY					
93	13	130	FADLMGYI					
79	11	130	FADLMGYIPL					
79	11	130	FADLMGYIPLV					
100	14	1927	FASGNMIV					
86	12	1927	FASGNMIVSPIT					
100	14	1773	FISGOYL	0.1000				
100	14	1773	FISGOYLA					
100	14	1773	FISGOYLAGL					
100	14	1304	HVDGGCSGGA					
79	11	177	FLALISCL	0.0046				
86	12	177	FLALISCLT					
86	12	177	FLALISCLT					
93	13	728	FLLLDANIV	0.2800	0.0480	0.0670	0.0150	0.3600
86	12	1228	FOVAHLIHA					
86	12	1228	FOVAHLIAPT					
79	11	2646	FOYSGOIN					
100	14	2792	FTEAMIRYSA					
93	13	1567	FTGLIHIDA					

ILCV A92 Super Motif with Binding Information

Conservancy	Freq	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
93	13	512	FTSPVWV					
93	13	512	FTSPVWGT					
93	13	512	FTSPVYGT					
79	11	684	FTLPALST					
79	11	684	FTLPALSTGL					
79	11	146	GAHAIALIGV					
86	12	982	GADTAACGDI					
86	12	992	GADTAACGDI					
86	12	1861	GAGVAGAL					
86	12	1861	GAGVAGALV					
86	12	1861	GAGVAGALVA					
86	12	350	GALVGV					
79	11	1895	GALVGVCA					
79	11	1895	GALVGVCA					
79	11	1885	GALVGVCA					
86	12	1345	GALVVI					
79	11	1345	GALVVIAT					
79	11	1345	GALVVIATA					
79	11	1345	GALVVIATAT					
100	14	1916	GAVQWNRIL					
100	14	1916	GAVQWNRIL					
100	14	1333	GIGTLDQA					
100	14	1333	GIGTLDQA					
100	14	1776	GIOYLAGL					
100	14	1776	GIOYLAGIST					
100	14	1776	GIOYLAGISTL					
79	11	1425	GIDVSVPT					
93	13	1552	GIPVQDIL					
79	11	968	GLHDLAVA					
79	11	968	GLHDLAVAV					
100	14	1782	GLSTUPGNPA					
79	11	1782	GLSTUPGNPA					
93	13	1569	GLTHDAHL					
93	13	28	GGACGM					
93	13	28	GONGGVIL					
93	13	2063	GTFPINAYT					
79	11	2063	GTFPINAYTT					
100	14	1335	GTVLDOAET					
100	14	1335	GTVLDOAETA					
86	12	1863	GVAGALVA					
79	11	1081	GVCWTVHGA					

HCV A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0208	A*6002
86	12	1870	GVLAALAA					
86	12	1670	GVLMIAAYCL					
79	11	161	GVNYATGNL	0.0001				
86	12	45	GVRAIHKI					
100	14	2619	GVTRCEKM					
100	14	2619	GVRCCKMA	0.0002				
100	14	2619	GVTRCEKML	0.0001				
83	13	154	GVRIEDGV					
79	11	1900	GVVCAIL					
100	14	1234	IIAPTSGKSI					
100	14	1572	HIDAHLSOT					
86	12	696	HIJKNNDV	0.0100	0.0014	0.5400	0.0027	0.0037
79	11	1719	HIJYIEQIM					
93	13	1769	HIANNFISG	0.3300	0.0004	0.1300	0.0280	0.0053
79	11	698	HIKNVDVOYL					
79	11	222	HIIPGCVPCV					
86	12	2855	HIIPVNSWL					
86	12	2855	IIIPVNSWLGM					
79	11	1910	IIMGRGGA					
86	12	1910	IIMGRGGA					
79	11	1910	IIMGRGGA					
86	12	1933	IIVSPTHIV					
100	14	1925	IIVSPTHIV	0.0430	0.0300	2.0000	0.0049	0.0450
79	11	1856	IIVSPTHIV	0.0002				
79	11	1656	ILAGYGAGVA					
86	12	1816	ILGGWMAA	0.0430	0.0024	0.0190	0.0005	0.0039
86	12	1816	ILGGWMAAQL					
86	12	1816	ILGGWMAAQL					
86	12	1331	ILGGTVL					
86	12	1331	ILGGTVL					
86	12	1331	ILGGTVL					
83	13	1891	ILSPGALV	0.0210	0.0004	0.3700	0.0036	0.0130
83	13	1891	ILSPGALV					
93	13	1891	ILSPGALV					
79	11	2591	IMAKNEVFCV	0.0088				
100	14	1777	IMAKNEVFCV					
100	14	1777	IMAKNEVFCV					
86	12	2250	ITRVESENKV					
86	12	2250	ITRVESENKV					
100	14	2816	ITSCSSNV					
100	14	2816	ITSCSSNV					
100	14	2816	ITSCSSNV					
86	12	909	ITWGADTA					
86	12	909	ITWGADTA					
86	12	909	ITWGADTA					

HCY A02 Super Motif with Binding Information

Conservancy	Freq	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
79	11	1296	ITYSTYGKFI					
79	11	1296	ITYSTYGKFI					
79	11	2613	WFPD LGV	0.0018				
79	11	2613	WFPD LGV					
93	13	30	MGAVYL					
86	12	1738	KALGILOT					
86	12	1736	KALGILOT					
86	12	2625	KMALYDVV					
86	12	1734	KOKAGIL					
86	12	1734	KOKAGILLOT					
86	12	1734	KOKAGILLOT					
86	12	121	KVIDITCGFA					
86	12	1255	KVILNP SV	0.0048				
100	14	1255	KVILNP SV					
100	14	1255	KVILNP SV					
100	14	1255	KVILNP SV					
79	11	1244	KVPAAVAA	0.0011				
86	12	1872	LAALAVCL					
79	11	1305	LADGGCGGA					
86	12	1729	LAOFKOKA					
86	12	1729	LAOFKOKAL					
79	11	1857	LAGYGAGV					
79	11	1857	LAGYGAGVA					
79	11	1857	LAGYGAGVAGA					
100	14	151	LAHGVRL					
86	12	179	LALISCLT					
79	11	972	LAVAEV					
100	14	1924	LAFASRGNIV	0.0004				
100	14	2815	LITSCSSNV					
100	14	2815	LITSCSSNVSV	0.0002				
79	11	2612	LWFPDLGV					
79	11	2612	LWFPDLGV					
86	12	178	LLALLSCL					
86	12	178	LLALLSCLT	0.0230	0.0150	0.0220	0.0011	0.0130
100	14	726	LEFLLADA					
93	13	726	LEFLLADARV	1.2000	0.0380	3.1000	0.1900	1.2000
86	12	1812	LLFNLGGWV					
86	12	1812	LLFNLGGWVA					
93	13	729	LLADARV	0.0061				
93	13	1887	LLPAILSPGA					
93	13	1887	LLPAILSPGAL	0.0025				
83	13	38	LLPRNGPHL					
93	13	36	LLPRNGPHLGV					

HCY A02 Super Motif with Binding Information

Conservancy	Freq	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6602
86	12	2240	LLWRQELMGKN					
83	13	1629	LLYRLGAV					
79	11	133	LMGYPLV					
79	11	133	LMGYPLVGA					
86	12	2761	LOOCTIAV					
88	12	126	LTCGFADL					
86	12	126	LTCGFADLM					
100	14	2180	LTDPSHIT					
100	14	2180	LTDPSHITA					
86	12	1052	LTRDKKOV					
86	13	1570	LTHIDAHFL					
93	13	2176	LTSMLTORSHI					
79	11	2738	LTTSCGNT					
79	11	2738	LTTSCGNL					
79	11	2738	LTTSCGNLIT					
86	12	1591	LVAYOATV					
86	12	1591	LVAYOATVCA	0.0002				
79	11	1853	LVDILAGYGA	-0.0001				
86	12	1667	LVGVLAA					
86	12	1667	LVGVLAAI	0.0003				
86	12	1667	LVGVLAAIA					
80	12	1667	LVGVLAAIAA					
100	14	1257	LVLPNSVA					
100	14	1257	LVLPNSVAA					
100	14	1257	LVLPNSVAAT					
100	14	1257	LVLPNSVAATL					
79	11	1864	LVMLPAIL					
79	11	1864	LVMLPAIL	0.0002				
79	11	1864	LVMLPAIL					
86	12	1137	LVTRIHADV					
79	11	1137	LVTRIHADVI	0.0001				
79	11	1137	LVTRIHADVIPV					
79	11	1897	LVGWGCA					
79	11	1897	LVGWGCAA					
79	11	1897	LVGWGCAAI	0.0011				
79	11	1897	LVGWGCAAIL					
79	11	1637	LVVIGESAI					
79	11	2773	LVVIGESA					
86	12	1348	LVVILATAT					
86	12	2592	MAKNEVECV	0.0022				
100	14	2179	MLTDPSHI					
100	14	2179	MLTDPSHIT	0.0002				
100	14	2179	MLTDPSHITA					
93	13	322	MMMNWSPIT					

HCY A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
93	13	1418	NAVAVYRGL					
93	13	1418	NAVAVYRGLDV					
86	12	2068	NAVTTGPCT					
86	12	1815	NILGWVA					
86	12	1815	NILGWVAA					
86	12	1815	NILGWVAAOL					
93	13	1282	NIRIGVAT	0.0001				
79	11	1282	NIRIGVHTI					
79	11	1282	NIRIGVHTIT					
79	11	1282	NIRIGVHTITTT					
86	12	2249	NITRVESENKV					
86	12	700	NIVDOYL					
86	12	118	NLGRVIDT					
86	12	118	NLGRVIDTL	0.0006				
86	12	118	NLGRVIDTLT					
93	13	1888	NLLPALSPGA					
86	12	2239	NLYMOEM	0.0041				
93	13	168	NIPGCSFSI					
93	13	168	NIPGCSFSRL					
86	12	1460	NTCVIOTV					
93	13	416	NINGSWH					
86	12	14	NTNIRVXDY					
93	13	1889	PALSPGA					
93	13	1889	PALSPGAL					
86	12	1889	PALSPGALV					
86	12	1889	PALSPGALVV					
86	12	688	PALSTGLI					
86	12	688	PALSTGLIHL					
79	11	2609	PARLWEPDL					
79	11	2066	PINAVTTGPCT					
79	11	1295	PITYSTYKFL					
79	13	2403	PLEGEPQYDL					
93	11	143	PLGGAARA	0.0001				
79	11	143	PLGGAARAL					
79	11	143	PLGGAARALA					
93	13	1628	PULYRLGA					
93	13	1628	PULYRLGAV	0.0001				
79	11	2667	PMGFSYDT					
79	11	2807	POPEYDEL					
79	11	2807	POPEYDELU					
79	11	2807	POPEYDELUIT					
93	13	7	PORKKXNIT					

HCV A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
86	12	109	PIIDRRNSIRL					
79	11	1473	PIETIETT					
79	11	1473	PIETIETT					
100	14	1236	PIGSGKST					
93	13	1236	PTGSGKSTIV					
86	12	1936	PIHYVPESDA					
86	12	1936	PIIIVPESDA					
79	11	1621	PIUGPTPL					
79	11	1621	PIUGPTPL					
78	11	2070	PTLWARMIL					
79	11	2870	PTLWARMIL					
79	11	2870	PTLWARMIL					
78	11	2870	PTLWARMIL					
100	14	1826	PTPLLYRL					
93	13	1826	PIPLYLGA					
93	13	1826	PIPLYLGA					
100	14	2857	PVNSWLGNI	0.0001				
100	14	2857	PVNSWLGNI	0.0001				
86	12	2857	PVNSWLGNI					
79	11	2318	PVNSWLGNI					
93	13	508	PVNSWLGNI					
93	13	508	PVNSWLGNI					
86	12	1340	OAETAGARL					
86	12	1340	OAETAGARL					
86	12	1340	OAETAGARL					
86	12	1340	OAETAGARL					
80	12	1603	OAETAGARL					
93	13	1595	OAETAGARL					
79	11	1595	OAETAGARL					
93	13	29	QATVCAIA					
93	13	29	QATVCAIA					
93	13	29	QATVCAIA					
86	12	336	QATVCAIA					
86	12	2164	QATVCAIA					
79	11	2210	QATVCAIA					
79	11	2210	QATVCAIA					
86	12	1465	QATVCAIA					
86	12	1465	QATVCAIA					
86	12	1465	QATVCAIA					
86	12	1465	QATVCAIA					
79	11	1186	QATVCAIA					
79	11	1186	QATVCAIA					
100	14	149	QATVCAIA					
100	14	149	QATVCAIA					
86	12	2733	QATVCAIA					
79	11	43	QATVCAIA					

HCY A02 Super Motif with Binding Information

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
78	11	2918	RLKLSAFL					
79	11	2611	RLWFDL	0.0280	0.0055	0.0180	0.0002	0.0032
79	11	2611	RLWFDL	0.0890	0.0110	1.0000	0.0100	0.0050
79	11	1618	RLKPTUIGPT					
86	12	1029	RLAPITA					
86	12	1347	RLVLATA					
86	12	1347	RLVLATAT					
100	14	619	RLWIPCT					
86	12	317	RLAWDMM					
93	13	635	RLWVGVEFL					
86	12	2243	RLKMGNI					
86	12	2243	RLKMGNT					
86	12	2243	RLKMGNTIV					
86	12	2243	RLKMGNTIV					
79	11	1284	RLGVHTIT					
79	11	1284	RLGVHTIT					
100	14	2621	RLCEKML					
86	12	2621	RLCEKMLYDV					
86	12	2252	RLSEKNV	0.0001				
86	12	2252	RLSEKNV					
79	11	2100	RLGDIIV					
86	12	156	RLLEDGVNA					
86	12	156	RLLEDGVNAT					
86	12	156	RLLEDGVNAT					
86	12	156	RLLEDGVNAT					
79	11	1655	RLVYLTHDPT					
79	11	1655	RLVYLTHDPT					
79	11	1655	RLVYLTHDPT					
79	11	1655	RLVYLTHDPT					
79	11	2212	RLVYLTHDPT					
79	11	2212	RLVYLTHDPT					
93	13	2207	RLVYLTHDPT					
100	14	175	RLVYLTHDPT					
86	12	175	RLVYLTHDPT					
86	12	175	RLVYLTHDPT					
100	14	1470	RLVYLTHDPT					
86	12	1470	RLVYLTHDPT					
86	12	1470	RLVYLTHDPT					
79	11	1470	RLVYLTHDPT					
79	11	1470	RLVYLTHDPT					
79	11	2926	RLVYLTHDPT	0.0008				
79	11	2926	RLVYLTHDPT	0.0002				
86	12	1051	RLVYLTHDPT	0.0053				
100	14	2178	RLVYLTHDPT					
100	14	2178	RLVYLTHDPT					
100	14	2178	RLVYLTHDPT					
86	12	2163	RLVYLTHDPT					
93	13	2209	RLVYLTHDPT					
79	11	2209	RLVYLTHDPT					
79	11	2209	RLVYLTHDPT					
79	11	2209	RLVYLTHDPT					

Conservancy	Freq.	Position	Sequence	A*0201	A*0202	A*0203	A*0208	A*6802
93	13	56	SQPKGRHPI					
86	12	1242	STKVPAAYA					
79	11	1242	STKVPAJYAA					
100	14	1784	STLPGNPA					
79	11	1784	STLPGNPAI					
79	11	2	STNPKPQIKT					
86	12	1663	SIWLVGGV					
86	12	1663	STWLVGGVL					
86	12	1663	STWLVGGVLA					
86	12	1299	STYGFIA					
100	14	1282	SVAILGFGA					
86	12	1455	SVDCNTCV					
86	12	1455	SVDCNTCVT					
86	12	995	TAACGDII					
86	12	1343	IAGAILVV					
86	12	1343	TAGARLVVL					
86	12	1343	TAGARLVILA					
86	12	1343	TAGARLVILAT					
79	11	1343	TARHLPVNSWL					
79	11	2852	TIMAKNEV					
79	11	2590	TICFGAYM					
93	13	1266	TICFGAYM					
86	12	1266	TICFGAYMSKA					
79	11	1622	TLHGPTPL					
79	11	1622	TLHGPTPL					
86	12	1811	TLFNLGGW					
86	12	1811	TUPALSTGL					
79	11	686	TUPALSTGL					
79	11	886	TUPALSTGL					
79	11	1785	TILPGNPAI					
79	11	125	TLICGFADL					
86	12	125	TLICGFADLM					
86	12	125	TLICGFADLM					
79	11	2871	TLWARMIL					
79	11	2871	TLWARMILM					
79	11	2871	TLWARMILMT					
86	12	1209	TMRSPIFT					
86	12	1464	TOYDFSI DPT					
86	12	1464	TOYDFSI DPT					
79	11	2589	TIIMAKNEV					
79	11	685	TIIPALST					
79	11	685	TIIPALSTGL					
79	11	685	TIIPALSTGLI					
86	12	1208	TMRSPIFT					
79	11	2739	TTSCGNIL					

HCY A02 Super Motif with Binding Information

Conservancy	Freq	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
79	11	2739	TTSCGNLIT					
79	11	1597	TVCAHADA					
86	12	1466	TVDFSLDPT					
86	12	1466	TVDFSLDPTFT					
100	14	1336	TVLDOAET					
100	14	1336	TVLDOAETA					
86	12	1336	TVLDOAETAGA					
100	14	1263	VAAITLGFGA					
93	13	1263	VAAITLGFAYM					
86	12	1230	VAHLHAPT					
86	12	1440	VATDALMT					
86	12	1592	VAYQATVCA	0.0005				
79	11	1592	VAYQATVCABA					
100	14	1420	VAYRGLDV	0.0001				
100	14	1420	VAYRGLDVS					
86	12	1456	VICNTCV					
86	12	1456	VICNTCVT					
86	12	1456	VICNTCVTOT					
86	12	1456	VIDILTCGFA					
86	12	1422	VLAALAYCL	0.0500	0.0087	0.0047	0.0002	0.0550
86	12	1671	VLCGYDA					
93	13	1521	VLCGYDACA					
79	11	1521	VLDQETA					
100	14	1337	VLDQETAGA					
86	12	157	VLEDGVNAT					
86	12	157	VLEDGVNAT					
100	14	1258	VLNPSVAA					
100	14	1258	VLNPSVAAT	0.0015				
100	14	1258	VLNPSVAATL					
100	14	1258	VLTTCGNIT					
79	11	2737	VLTTCGNIT	0.0002				
79	11	2737	VLTTCGNITL					
79	11	2737	VLTTCGNITL					
79	11	1852	VLDLAGYGA					
86	12	1666	VLVGVLA					
86	12	1666	VLVGVLA	0.0270	0.0130	0.3100	0.0120	0.0130
86	12	1666	VLVGVLAAL	0.0084				
86	12	1666	VLVGVLAAL					
100	14	1256	VLVNPSV	0.0009				
100	14	1256	VLVNPSVA					
100	14	1256	VLVNPSVAA					
100	14	1256	VLVNPSVMA					
100	14	1256	VLEKGNKTA					
79	11	2600						

ILCV A02 Super Motif with Binding Information

Conservancy	Freq	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
100	14	1918	VQMMNRRL					
100	14	1918	VQMMNRRLA					
100	14	1918	VQMMNRRLAFA					
86	12	1463	VTQIVDFSL					
79	11	1138	VTRHADV					
79	11	1138	VTRHADVIV					
86	12	1661	VTSTWLV					
86	12	1661	VTSTWLVGCV					
79	11	1439	VVA1DALM					
79	11	1439	VVA1DALMT					
79	11	1901	VVCALLRIIV					
79	11	1898	VGVVCAAI					
79	11	1898	VGVVCAAIL					
86	12	1660	VV1STWVL					
86	12	1660	VV1STWVLV	0.0003				
88	12	1766	WAKLMMNEI	0.0001				
86	12	76	WAPGYWPL					
06	12	2873	WARMILMT					
79	11	2287	WARPDPNPL					
100	14	1920	WMNRLLAFA	0.0410	0.0330	3.0000	0.0023	0.1000
79	11	557	WMNSTGT					
86	12	1665	WVLVGGVL					
86	12	1665	WVLVGGVLA	0.0005				
86	12	1665	WVLVGGVLAAL	0.0015				
86	12	1665	WVLVGGVLAAL					
79	11	1249	YAAOGYKV					
79	11	1249	YAAOGYKVL					
79	11	1249	YAAOGYKVLV					
79	11	1249	YAAOGYKVLV					
79	11	136	YPLVGAFL	0.0050				
100	14	1779	YLAGLSTL					
86	12	1185	YLGSSGGPL	0.0002				
86	12	1185	YLGSSGGPL					
93	13	35	YLTTRGCTIL	0.0400	0.0007	0.0220	0.0089	0.0039
79	11	2836	YLTTRPIT					
86	12	1580	YLVAYOAT					
86	12	1590	YLVAYOATV	0.2500	0.1100	0.6300	0.0450	1.2000
86	12	1590	YLVAYOATVCA					
86	12	1138	YLVIRHADV	0.0110	0.0021	2.8000	0.0520	0.0130
79	11	1136	YLVIRHADV					
93	13	1594	YQATVCAFLA					

HCY A02 Super Motif with Binding Information

Conservancy	Freq	Position	Sequence	A*0201	A*0202	A*0203	A*0206	A*6802
79	11	1594	YOATVCAHAQA					
79	11	1106	YIMDDDL					
79	11	1106	YIMDDDLV					
86	12	276	YVGLCGSV	0.0018				
86	12	276	YVGLCGSVFL					
93	13	637	YVGSVEHFL	0.0008				
86	12	1939	YVPESDAA					
88	12	1939	YVPESDAAA					
86	12	1939	YVPESDAANIV					
			555					

HCY A03 Super Motif (With Binding Information)

119

HCY A03 Super Motif (With Binding Information)

Conservancy	Freq	Position	Sequence	A*0301	A*1101	A*3101	A*3301	A*6801
86	12	2235	LIEAILWR	0.0008	0.0005	0.0018	0.0069	0.0008
100	14	1398	LIFCHSKK	0.5400	0.1900	0.0071	0.0012	0.0240
100	14	1396	LIFCHSKK					
79	11	2612	LWFDLGVH	0.0003	0.0001			
100	14	726	LLFLLADAR					
93	13	36	LLPRKGP1					
86	12	87	LLSPKGS1					
79	11	1591	LVAYOATVCAR					
79	11	1	MSINPKPQ1H					
79	11	1	MSINPKPQ1H					
86	12	2248	NITRVESENK	0.0010	0.0062			
79	11	14	NINRPQDYK	0.0010	0.0007			
79	11	1295	PITYSTYK					
79	11	2667	PMGESYD1H					
83	13	514	PSPVAVG1D1H					
79	11	1607	PSMOCKMMK					
86	12	109	PTDPRR1HSA	0.0008	0.0005	0.0006	0.0006	0.0002
93	13	1236	PTGSGKSTK	0.0002	0.0001			
93	13	616	PVAVG1D1H	0.0008	0.0005			
86	12	1340	QAE1AGAR					
93	13	28	QMGVYL1PH					
86	12	289	QLFTSPR					
79	11	289	QLFTSPR1	0.7500	0.0330	0.0290	0.0077	3.1000
79	11	2210	OKSAPSLK					
100	11	1188	RAAVCTRGVAK					
79	14	149	RALAIGVR					
79	11	47	RATIKTSEH					
79	11	43	RLGVRA1H					
79	11	43	RLGVRA1H	0.9400	0.0290	0.0420	0.0004	0.0001
100	14	1923	RLVEASR					
79	11	2611	RLWFDLGVH					
100	14	636	RMVVGVE1H	0.7200	0.0200	0.1900	0.0030	0.0045
93	13	55	FSQTKGP1H					
79	11	2207	SASQLSAPSLK					
86	12	1132	SSDLV1VR	0.0003	0.0044			
79	11	2	STNPKPQ1					
79	11	2	STNPKPQ1					
79	11	2	STNPKPQ1KTK					
86	12	1266	TLGFGAMSK	0.0810	0.0610	0.0005	0.0013	0.0009
79	11	1622	TUHGPTLLYH					
93	13	52	TSENSQPH					
86	12	52	TSENSQPH	0.0003	0.0001			
86	12	52	TSENSQPH					
86	12	1050	TSLTGROK					
86	12	1864	VAGALVAFK	0.2400	0.8900	0.0048	0.0025	0.0310
79	11	1592	VAYOATVCAR	0.0005	0.0038	0.0680	0.0720	0.0280
86	12	1337	VLDQAE1AGAR					
79	11	1138	VTR1ADVPVR					
79	11	1901	VWCAAILR					
79	11	1901	VWCAAILR					
79	11	1888	WGVWCAAILR					
93	13	517	WVG1D1H					

HCY A93 Super Motif (With Binding Information)

Conservancy	Freq	Position	Sequence	A*0301	A*1101	A*3101	A*3301	A*6801
88	12	93	WAGMLSPR					
88	12	96	WLSPTGCSR	0.0008	0.0005			
100	14	1920	WMANRLIAFASR					
79	11	557	WMNSTIGFTK	0.0530	0.0810	0.0014	0.0420	0.0056
93	13	35	YLPFRGSPR	0.0054	0.0005			
79	11	2930	YSPGEINL					
100	14	637	YVGGEINL					
86	12	1939	YVESDAAR	0.0003	0.0001			
		112						

Table X

HCY A24 Super Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*2401
ALLSPGAL	1890	8	13	93	
ALAHGVHL	150	9	14	100	
ALSTGILHL	689	9	12	86	
ALVGVVCAAI	1896	11	11	79	
ATGILPGCSF	165	10	13	93	
ATIIGI GAY	1265	5	12	100	
ATIIGI GAY	1265	9	13	93	
AVAVYHGL	1419	8	14	100	
AVQWNRIL	1917	8	14	100	
AVQWNRIL	1917	9	14	100	
AVQWNRILIAF	1917	11	14	100	
AWDMNMNW	319	8	12	86	0.0009
AVAAQGYKL	1248	10	11	79	
AVYRGLDVSVI	1421	11	14	100	
CIHKLGVPL	2941	10	12	86	
CIWMMLI	739	8	12	86	
CIQGSSDL	1128	8	11	79	
CIQGSSDL	1128	9	11	79	0.0001
CIQGSSDL	1128	10	11	79	
CIQGSSDL	1128	11	11	79	
CIQGSSDL	1190	11	11	79	
CIQGSSDL	1190	9	11	79	
CIQGSSDL	1190	8	11	79	
CIQGSSDL	1190	10	11	79	
CIQGSSDL	1190	11	11	79	
CIQGSSDL	1190	12	11	79	
CIQGSSDL	1190	13	11	79	
CIQGSSDL	1190	14	11	79	
CIQGSSDL	1190	15	11	79	
CIQGSSDL	1190	16	11	79	
CIQGSSDL	1190	17	11	79	
CIQGSSDL	1190	18	11	79	
CIQGSSDL	1190	19	11	79	
CIQGSSDL	1190	20	11	79	
CIQGSSDL	1190	21	11	79	
CIQGSSDL	1190	22	11	79	
CIQGSSDL	1190	23	11	79	
CIQGSSDL	1190	24	11	79	
CIQGSSDL	1190	25	11	79	
CIQGSSDL	1190	26	11	79	
CIQGSSDL	1190	27	11	79	
CIQGSSDL	1190	28	11	79	
CIQGSSDL	1190	29	11	79	
CIQGSSDL	1190	30	11	79	
CIQGSSDL	1190	31	11	79	
CIQGSSDL	1190	32	11	79	
CIQGSSDL	1190	33	11	79	
CIQGSSDL	1190	34	11	79	
CIQGSSDL	1190	35	11	79	
CIQGSSDL	1190	36	11	79	
CIQGSSDL	1190	37	11	79	
CIQGSSDL	1190	38	11	79	
CIQGSSDL	1190	39	11	79	
CIQGSSDL	1190	40	11	79	
CIQGSSDL	1190	41	11	79	
CIQGSSDL	1190	42	11	79	
CIQGSSDL	1190	43	11	79	
CIQGSSDL	1190	44	11	79	
CIQGSSDL	1190	45	11	79	
CIQGSSDL	1190	46	11	79	
CIQGSSDL	1190	47	11	79	
CIQGSSDL	1190	48	11	79	
CIQGSSDL	1190	49	11	79	
CIQGSSDL	1190	50	11	79	
CIQGSSDL	1190	51	11	79	
CIQGSSDL	1190	52	11	79	
CIQGSSDL	1190	53	11	79	
CIQGSSDL	1190	54	11	79	
CIQGSSDL	1190	55	11	79	
CIQGSSDL	1190	56	11	79	
CIQGSSDL	1190	57	11	79	
CIQGSSDL	1190	58	11	79	
CIQGSSDL	1190	59	11	79	
CIQGSSDL	1190	60	11	79	
CIQGSSDL	1190	61	11	79	
CIQGSSDL	1190	62	11	79	
CIQGSSDL	1190	63	11	79	
CIQGSSDL	1190	64	11	79	
CIQGSSDL	1190	65	11	79	
CIQGSSDL	1190	66	11	79	
CIQGSSDL	1190	67	11	79	
CIQGSSDL	1190	68	11	79	
CIQGSSDL	1190	69	11	79	
CIQGSSDL	1190	70	11	79	
CIQGSSDL	1190	71	11	79	
CIQGSSDL	1190	72	11	79	
CIQGSSDL	1190	73	11	79	
CIQGSSDL	1190	74	11	79	
CIQGSSDL	1190	75	11	79	
CIQGSSDL	1190	76	11	79	
CIQGSSDL	1190	77	11	79	
CIQGSSDL	1190	78	11	79	
CIQGSSDL	1190	79	11	79	
CIQGSSDL	1190	80	11	79	
CIQGSSDL	1190	81	11	79	
CIQGSSDL	1190	82	11	79	
CIQGSSDL	1190	83	11	79	
CIQGSSDL	1190	84	11	79	
CIQGSSDL	1190	85	11	79	
CIQGSSDL	1190	86	11	79	
CIQGSSDL	1190	87	11	79	
CIQGSSDL	1190	88	11	79	
CIQGSSDL	1190	89	11	79	
CIQGSSDL	1190	90	11	79	
CIQGSSDL	1190	91	11	79	
CIQGSSDL	1190	92	11	79	
CIQGSSDL	1190	93	11	79	
CIQGSSDL	1190	94	11	79	
CIQGSSDL	1190	95	11	79	
CIQGSSDL	1190	96	11	79	
CIQGSSDL	1190	97	11	79	
CIQGSSDL	1190	98	11	79	
CIQGSSDL	1190	99	11	79	
CIQGSSDL	1190	100	11	79	
CIQGSSDL	1190	101	11	79	
CIQGSSDL	1190	102	11	79	
CIQGSSDL	1190	103	11	79	
CIQGSSDL	1190	104	11	79	
CIQGSSDL	1190	105	11	79	
CIQGSSDL	1190	106	11	79	
CIQGSSDL	1190	107	11	79	
CIQGSSDL	1190	108	11	79	
CIQGSSDL	1190	109	11	79	
CIQGSSDL	1190	110	11	79	
CIQGSSDL	1190	111	11	79	
CIQGSSDL	1190	112	11	79	
CIQGSSDL	1190	113	11	79	
CIQGSSDL	1190	114	11	79	
CIQGSSDL	1190	115	11	79	
CIQGSSDL	1190	116	11	79	
CIQGSSDL	1190	117	11	79	
CIQGSSDL	1190	118	11	79	
CIQGSSDL	1190	119	11	79	
CIQGSSDL	1190	120	11	79	
CIQGSSDL	1190	121	11	79	
CIQGSSDL	1190	122	11	79	
CIQGSSDL	1190	123	11	79	
CIQGSSDL	1190	124	11	79	
CIQGSSDL	1190	125	11	79	
CIQGSSDL	1190	126	11	79	
CIQGSSDL	1190	127	11	79	
CIQGSSDL	1190	128	11	79	
CIQGSSDL	1190	129	11	79	
CIQGSSDL	1190	130	11	79	
CIQGSSDL	1190	131	11	79	
CIQGSSDL	1190	132	11	79	
CIQGSSDL	1190	133	11	79	
CIQGSSDL	1190	134	11	79	
CIQGSSDL	1190	135	11	79	
CIQGSSDL	1190	136	11	79	
CIQGSSDL	1190	137	11	79	
CIQGSSDL	1190	138	11	79	
CIQGSSDL	1190	139	11	79	
CIQGSSDL	1190	140	11	79	
CIQGSSDL	1190	141	11	79	
CIQGSSDL	1190	142	11	79	
CIQGSSDL	1190	143	11	79	
CIQGSSDL	1190	144	11	79	
CIQGSSDL	1190	145	11	79	
CIQGSSDL	1190	146	11	79	
CIQGSSDL	1190	147	11	79	
CIQGSSDL	1190	148	11	79	
CIQGSSDL	1190	149	11	79	
CIQGSSDL	1190	150	11	79	
CIQGSSDL	1190	151	11	79	
CIQGSSDL	1190	152	11	79	
CIQGSSDL	1190	153	11	79	
CIQGSSDL	1190	154	11	79	
CIQGSSDL	1190	155	11	79	
CIQGSSDL	1190	156	11	79	
CIQGSSDL	1190	157	11	79	
CIQGSSDL	1190	158	11	79	
CIQGSSDL	1190	159	11	79	
CIQGSSDL	1190	160	11	79	
CIQGSSDL	1190	161	11	79	
CIQGSSDL	1190	162	11	79	
CIQGSSDL	1190	163	11	79	
CIQGSSDL	1190	164	11	79	
CIQGSSDL	1190	165	11	79	
CIQGSSDL	1190	166	11	79	
CIQGSSDL	1190	167	11	79	
CIQGSSDL	1190	168	11	79	
CIQGSSDL	1190	169	11	79	
CIQGSSDL	1190	170	11	79	
CIQGSSDL	1190	171	11	79	
CIQGSSDL	1190	172	11	79	
CIQGSSDL	1190	173	11	79	
CIQGSSDL	1190	174	11	79	
CIQGSSDL	1190	175	11	79	
CIQGSSDL	1190	176	11	79	
CIQGSSDL	1190	177	11	79	
CIQGSSDL	1190	178	11	79	
CIQGSSDL	1190	179	11	79	
CIQGSSDL	1190	180	11	79	
CIQGSSDL	1190	181	11	79	
CIQGSSDL	1190	182	11	79	
CIQGSSDL	1190	183	11	79	
CIQGSSDL	1190	184	11	79	
CIQGSSDL	1190	185	11	79	
CIQGSSDL	1190	186	11	79	
CIQGSSDL	1190	187	11	79	
CIQGSSDL	1190	188	11	79	
CIQGSSDL	1190	189	11	79	
CIQGSSDL	1190	190	11	79	
CIQGSSDL	1190	191	11	79	
CIQGSSDL	1190	192	11	79	
CIQGSSDL	1190	193	11	79	
CIQGSSDL	1190	194	11	79	
CIQGSSDL	1190	195	11	79	
CIQGSSDL	1190	196	11	79	
CIQGSSDL	1190	197	11	79	
CIQGSSDL	1190	198	11	79	
CIQGSSDL	1190	199	11	79	
CIQGSSDL	1190	200	11	79	
CIQGSSDL	1190	201	11	79	
CIQGSSDL	1190	202	11	79	
CIQGSSDL	1190	203	11	79	
CIQGSSDL	1190	204	11	79	
CIQGSSDL	1190	205	11	79	
CIQGSSDL	1190	206	11	79	
CIQGSSDL	1190	207	11	79	
CIQGSSDL	1190	208	11	79	
CIQGSSDL	1190	209	11	79	
CIQGSSDL	1190	210	11	79	
CIQGSSDL	1190	211	11	79	
CIQGSSDL	1190	212	11	79	
CIQGSSDL	1190	213	11	79	
CIQGSSDL	1190	214	11	79	
CIQGSSDL	1190	215	11	79	
CIQGSSDL	1190	216	11	79	
CIQGSSDL	1190	217	11	79	
CIQGSSDL	1190	218	11	79	
CIQGSSDL	1190	219	11	79	
CIQGSSDL	1190	220	11	79	
CIQGSSDL	1190	221	11	79	
CIQGSSDL	1190	222	11	79	
CIQGSSDL	1190	223	11	79	
CIQGSSDL	1190	224	11	79	
CIQGSSDL	1190	225	11	79	
CIQGSSDL	1190	226	11	79	
CIQGSSDL	1190	227	11	79	
CIQGSSDL	1190	228	11	79	
CIQGSSDL	1190	229	11	79	
CIQGSSDL	1190	230	11	79	
CIQGSSDL	1190	231	11	79	
CIQGSSDL	1190	232			

123

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A-2401
FISGKQL	1773	8	14	100	
FISGIQYAGL	1773	11	14	100	
FLIALLSQL	177	9	12	86	
FTEAMITY	2792	8	14	100	
FTGLTHDAIF	1567	11	13	93	
FTLLPALSTGL	bv4	11	11	79	
FWAKWMMNF	1765	9	12	86	6.9000
FVNAQIMMNF	1765	10	12	86	
GFADUMGY	129	8	13	93	
GFADLMGYI	129	9	13	93	
GFADLAGVPL	129	11	11	79	
GFSYDTDF	2689	9	11	79	
GIONTAQL	1776	8	14	100	
GIONTAGLSTL	1776	11	14	100	
GLPYCOXHL	1552	0	13	100	
GLPYCOXHLEF	1552	11	12	93	
GLSAFSLSLY	2921	10	11	86	
GLSTLPGNPN	1782	11	11	79	0.0001
GLTHIDAF	1589	9	11	79	
GLTHIDAFIL	1569	10	13	93	
GLTFPINAY	2063	9	11	93	
GVAGALVAF	1863	8	12	79	
GVAKAVDF	1193	8	11	86	
GVLAALAAY	1670	9	12	79	
GVLAALAAYCL	1670	11	12	86	
GWNTATGL	161	9	11	86	
GWNRCEKA	2619	8	14	79	
GWNRCEKMVL	2619	10	14	100	
GWNRCEKNALY	2618	11	14	100	
GWNRCEKNALY	2618	11	14	100	
GWNRLEDQANT	154	11	12	100	
GVVCMAIL	1800	8	11	88	
GWRLIAP	1027	8	11	79	
GWRLIAPTAY	1027	11	11	79	
GYGAGVAGAL	1859	10	12	79	0.0003
GYPLVGAPL	135	10	11	86	
GYRRCRASGL	2728	11	12	79	0.0057
HLDNDVDVOY	696	11	12	86	
HLPPIECGA	1719	9	11	79	
HMMNFECSJ	1769	9	11	79	
HMMNFECSJOY	1789	11	13	93	
HTPNSML	2855	8	12	93	
HTPNSMLGINI	2855	11	12	86	
HMRGGEGANQM	1910	11	11	86	
IPLALLSQL	176	10	12	86	
ILCGWMAAQ	1816	10	12	86	0.0026

ICCV A24 Super Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*2401
ILGIGTVL	1331	8	12	86	
IMAKNEVF	2581	8	12	86	
ITYSTYGF	1296	9	12	86	
ITYSTYKTL	1296	10	11	86	
MDVQVLY	701	8	12	79	
NGGVYL	30	8	13	86	
KIPGQGI	23	8	13	93	
KVIDLTGCF	121	10	12	93	
LFNLGGW	1813	8	12	86	
LEANLW	2235	8	12	86	
LINNGSW	414	8	11	86	
LIALLSCL	170	8	12	79	
LLAPITAY	1030	8	14	86	
LEFLNGW	1812	9	12	100	
LLPAILSPGAL	1887	11	13	86	
LLPRGRL	36	9	13	93	
LLSPRGSPSW	97	11	11	93	
LLWROKNGW	2240	11	12	79	
LTCGFADL	126	8	12	86	
LTCGFADLM	126	9	12	86	
LTOGFADLMGY	126	11	12	86	
LTHIDAHF	1570	8	13	86	
LTHIDAFEL	1570	9	13	93	
LTSMLTDPWH	2178	11	13	93	
LITSCGNTL	2738	9	11	83	
LVDLIAGY	1853	8	11	79	
LVGGVLAAL	1667	9	12	79	
LVNIPSVAAITL	1257	11	14	86	
LVNLPAL	1804	8	11	100	
LVNLPAIL	1884	9	11	79	
LVTRIADVI	1137	9	11	79	
LVGWCAAI	1897	10	11	79	
LVGVCAAIL	1897	11	11	79	
LVARMLLM	2872	8	12	86	
LVARMLMTHF	2872	11	12	86	
LWROKNGW	2241	10	12	86	
LYLVTRHADV	1135	11	12	86	
MLMTHFF	2876	8	11	79	
MLTDPWH	2179	8	12	86	
MMNFSGI	1770	8	14	100	
MMNFSGIQY	1770	10	14	100	
MMNFSGIQYL	1770	11	14	100	
MYGVGVHRL	636	10	13	100	
NFSGIQY	1772	8	14	93	0.0270
NFSGIQYL	1772	9	14	100	0.0170

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*240
NILGGWAAQ	1815	11	12	86	
NIRIGVRI	1282	9	11	79	
NIVDOYL	700	8	12	86	
NIVDOYLY	700	9	12	86	
NILKVIDTL	118	9	12	86	
NILWYKDM	2232	8	12	86	0.0001
NILPGCSFSI	168	9	13	86	
NILPGCSFSI	168	10	13	93	
NILPGCSFSI	168	11	13	93	
NITCVQIVDF	1460	10	12	93	
NINSGMII	146	8	13	86	
NINRVDVWF	14	11	13	93	
NMDOLVGW	1108	9	11	79	
NMFCTVMA	551	8	11	79	
PIITYSYGKE	1295	10	12	86	
PIITYSYGKE	1295	11	11	79	
PIEGEGDPL	2403	11	11	79	
PIGGAAAL	143	9	13	93	
PMGFSYDTRCF	2667	11	11	79	
PIDRRNRNPL	109	11	11	79	
PIIHGPTPL	1621	9	12	86	
PIIHGPTPL	1621	10	11	79	
PIIHGPTPL	1621	11	11	79	
PTLWARM	2870	8	11	79	
PTLWARMIL	2870	9	11	79	
PTLWARMIL	2870	10	11	79	
PTPLVYRL	1626	8	14	100	
PVODDIEF	1554	9	12	86	
PVODDILFW	1554	10	12	86	
PVNSMIGNI	2867	8	14	100	
PVNSMIGNI	2867	10	14	100	
PVNSMIGNI	2867	11	12	86	
PVNSMIGNI	2867	12	11	79	
OROKAQL	1732	9	12	86	
OROKAQL	1732	10	12	86	
OWGGVYL	29	8	13	93	
OWGGVYL	29	9	13	93	
QIVDFSLDPTF	1465	11	12	86	
QVAGLSTL	1919	9	14	100	
QVAGLSTL	1919	10	14	100	
QVSGQVME	2647	10	11	79	
QVSGQVME	2647	11	11	79	
RLHQSAFSL	2918	8	12	86	
RLHQSAFSL	2918	10	11	79	
RLHFPL	2611	8	11	79	0.0001

Sequence	Position	Pepide No	No of Amino Acids	Sequence Frequency	Conservancy (%)	A*2401
RLLA11AY	1029		9	12	86	
RLAWDMNM	317		8	12	86	
RLAWDMNMW	317		10	12	86	
RLMLMTIF	2875		8	12	86	
RLMLMTIFF	2975		9	12	86	
RLMYGVDFRL	433		11	12	86	
RLVCEKML	2521		8	13	93	
RLVCEKMLY	2521		9	14	100	
RLVEDVNY	156		9	14	100	
RLSFLAL	173		9	12	86	
RLSFLALL	173		9	14	86	
RLSFLALL	173		10	14	100	
RLSFLALL	175		8	14	100	0.0041
RLSFLALL	175		11	14	100	
RLSFLALL	170		8	12	86	
RLSFLALL	1470		10	14	100	
RLSFLALL	2926		10	14	100	
RLSFLALL	2178		9	14	79	
RLSFLALL	1242		8	14	100	
RLSFLALL	1784		9	12	86	
RLSFLALL	1653		10	11	79	
RLSFLALL	1232		8	12	86	
RLSFLALL	1232		11	14	100	
RLSFLALL	1262		11	14	100	
RLSFLALL	1608		9	11	79	
RLSFLALL	2860		8	12	86	
RLSFLALL	1164		11	12	86	
RLSFLALL	2590		9	11	86	
RLSFLALL	1256		8	13	79	
RLSFLALL	1622		8	11	93	
RLSFLALL	1622		9	11	78	
RLSFLALL	1622		10	11	79	
RLSFLALL	1811		10	11	79	0.0001
RLSFLALL	1811		10	11	79	
RLSFLALL	686		9	12	80	
RLSFLALL	686		10	11	79	
RLSFLALL	1785		8	11	79	
RLSFLALL	125		9	11	79	
RLSFLALL	125		10	12	86	
RLSFLALL	2871		8	12	86	
RLSFLALL	2871		9	11	79	
RLSFLALL	2871		10	11	79	
RLSFLALL	2589		10	11	79	
RLSFLALL	685		10	11	79	
RLSFLALL	685		11	11	79	
RLSFLALL	1208		8	12	79	
RLSFLALL	2739		8	12	86	
RLSFLALL	1466		10	11	79	
RLSFLALL	556		8	12	86	
RLSFLALL	1664		9	11	79	
RLSFLALL	1664		12	12	86	

ICCV A24 Super Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*2401
TYSTYGKF	1297	8	13	93	
TYSTYGKL	1297	9	12	86	
VTGLTH	1566	8	13	93	0.0230
VIDLTGCF	122	9	12	86	
VLAALAAY	1871	8	12	86	
VLAALAAYCL	1671	10	12	86	0.0070
VLEDGMY	157	8	12	86	
VLPNSVAAIL	1258	10	14	86	
VLTTSCGNTL	2737	10	11	100	
VLVDLAGY	1852	9	11	79	
VLVGGVLAAL	1666	10	12	79	
VMGSSYGF	2039	0	11	86	
VMGSSYGFQY	2639	10	11	79	
VTQIVDFSL	1463	9	12	79	
VIRIADVI	1138	8	11	86	
VVAIDALM	1439	8	11	79	
VGVVCAAN	1898	9	11	79	
VGVVCAAIL	1890	10	11	79	
VTSITWIL	1660	8	12	79	
VVLPRNGPRL	34	11	13	86	
VMNRILAF	1920	8	14	83	0.0016
WMLVSGVL	1665	8	12	100	
WMLVGGVLAAL	1665	11	12	86	
YPLVGAPL	136	9	11	86	
YLAGLSTL	1779	8	14	79	
YKGSQSGRL	1165	10	12	100	
YKGSQSGRL	1165	11	12	86	
YLPPIRGPTL	35	10	13	86	
YLVTRHADV	1136	10	11	93	0.0001
YTNWDOQL	1106	8	11	79	
YTNWDOQLVGV	1106	11	11	79	
YWGDLQSSVIF	276	10	12	79	
YWGDLQSSVRL	276	11	12	86	
YWGVAEFTL	637	9	13	86	
YTRGLDVSIV	1422	10	14	93	
260		3		100	

Table XI

ILCV 107 Super Model (with Binding Information)

Consistency	Freq	Position	Sequence	B*0702	B*3501	B*5101	B*5301	B*5401
86	12	1604	APPSWDOM	0.0028	0.0002	0.0002	0.0001	0.0002
79	11	1604	APPSWDOMW	0.0001	0.0001	0.0002	0.0006	0.0003
93	13	1235	APTSCKSTKV	0.0001				
79	11	2869	APTLWARM	0.4300	0.0001	0.0012	-0.0002	0.0023
79	11	2869	APTLWARM	0.0160	0.0002	0.0012	0.0001	0.0002
79	11	2869	APTLWARM	0.8800	0.0001	0.0010	0.0001	0.0003
79	11	2869	APTLWARM	0.0130	0.0001	-0.0003	-0.0002	0.0033
79	11	2869	APTLWARM	0.0001	0.0002	0.0002	0.0005	0.0002
79	11	2410	APTLWARM	0.0001	0.0002	0.0001	0.0001	0.0002
86	12	111	DPHRSRL	0.0170	0.0002			
79	11	2615	FPDLGVRV	0.0001				
100	14	24	FTGGQGV	0.0001				
100	14	24	FTGGQGV	0.0001				
86	12	1912	GTGGQGV	0.0001	0.0002	0.0002	0.0001	0.0002
86	12	1912	GTGGQGV	0.0001	0.0001	0.0002	0.0001	0.0003
86	12	1912	GTGGQGV	0.0001	0.0001	0.0002	0.0001	0.0003
93	13	41	GPIIYRL	0.0001				
100	14	1625	GPIIYRL	0.0024	0.0002	0.0002	0.0001	0.0002
93	13	1625	GPIIYRL	0.0005				
93	13	507	GPYCHTSPV	0.0001				
83	13	1378	IPFYGKAI	0.0120	0.0001	0.1200	-0.0002	0.2000
79	11	137	IPYVGAFL	0.4400	0.0032	0.0700	0.0003	0.0035
86	12	2608	KPARLIV	0.0150	0.0002	0.0017	-0.0002	0.0008
79	11	2608	KPARLIV	0.0003				
79	11	2608	KPARLIV	1.4150	0.0001	0.0002	0.0001	0.0003
79	11	1820	KPTLHPTPL	0.0021				
93	13	1820	KPTLHPTPL	0.0001	0.0001	0.0001	0.0002	0.9400
93	13	1888	LPALSPGA	0.0053	0.0001	0.0036	0.0001	0.2100
93	13	1888	LPALSPGA	0.0003				
86	12	1808	LPALSPGALV	0.0020				
100	14	807	LPALSTGL	0.0350	0.0002	2.0000	0.0062	0.0005
86	12	887	LPALSTGL	0.0011				
86	12	887	LPALSTGL	0.0001	0.0002	0.0001	0.0001	0.0002
86	12	2165	LPCEPFDV	0.0110	0.0360	0.0059	0.0150	0.0018
93	13	109	LPCCSFSI	0.1950	0.0796	0.0550	0.0013	0.0015
93	13	169	LPCCSFSI	0.0022	0.0009	0.0100	0.0140	0.0012
93	13	169	LPCCSFSI	0.0007				
93	13	169	LPCCSFSI	6.5000	0.0001	0.0180	-0.0002	0.0020
93	13	37	LPKGPRL	0.1900	0.0001	0.0009	0.0001	0.0025
93	13	37	LPKGPRL	0.0005				
93	13	1553	LPVODHL	0.0001	0.0046	0.0002	0.0110	0.0003
86	12	1553	LPVODHL	0.0001				
86	12	1553	LPVODHL	0.0001	0.0001	0.0040	-0.0002	0.0013
86	12	1720	LPVODHL	0.0130	0.0001			
100	14	1260	NPVAVATL	0.0011	0.0001	0.0002	0.0001	0.0003
100	14	1260	NPVAVATL	0.0001	0.0001	0.0002	0.0001	0.0002
86	12	1605	PPSWDOM	0.0003				
79	11	1605	PPSWDOM	0.0001	0.0002	0.0001	0.0001	0.0002
79	11	1605	PPSWDOM	0.0002				
79	11	1608	PPSWDOM	0.0001				
79	11	1608	PPSWDOM	0.0001	0.0001	0.0001	0.0001	-0.0002
79	11	2317	PPWKGQFL	0.0140	0.0001	0.0001	0.0002	0.0180
79	11	2601	PPWKGQFL	0.0011	0.0001	0.0001	0.0002	0.0002
79	11	2808	PPWKGQFL	0.0002				
79	11	2808	PPWKGQFL	0.0001	0.0002	0.0002	0.0001	0.0002
86	12	78	PPWKGQFL	0.0006				

HCY B07 Super Motif (with Binding Information)

Conservancy	Freq	Position	Sequence	B*0702	B*3501	B*5101	B*5301	B*5401
86	12	78	OPGYMPLY	0.0001	0.0011	0.0002	0.0001	0.0002
83	13	57	ORXGRQRA	0.2300	0.0002	0.0001	0.0001	0.0002
79	11	2299	RPDYNPPL	0.0050				
93	13	1893	SPGALVGV	0.0001	0.0002	0.0002	0.1200	0.0002
79	11	1893	SPGALVGV	0.0130	0.0001	0.0016	0.0001	0.0003
79	11	2931	SPGENIV	0.0007				
79	11	2931	SPGEINVA	0.0003	0.0001	0.0001	0.0002	0.0037
79	11	2649	SPQRNEF	0.0027				
79	11	2649	SPQRNEF	0.1200	0.0002	0.0002	0.0001	0.0002
79	11	2649	SPRGRPSW	0.3800	0.0002	0.0005	0.0001	0.0002
86	12	99	SPTIYVPESDA	0.0001				
86	12	1935	TPCGSM	0.0028				
86	12	1975	TPCTGSSDL	0.0005	0.0001	0.0002	0.0001	0.0003
79	11	1126	TPCTGSSDL	0.0001				
79	11	1126	TPGCVPCV	0.0001				
86	12	223	TPGAPVODL	0.0001				
93	13	1550	TPLYRILGA	0.0083	0.0001	0.0001	0.0002	0.2300
93	13	1627	TPLYRILGA	0.0120	0.0001	0.0008	0.0001	0.0110
93	13	1827	TPVNSMGIN	0.0001	0.0001	0.0053	0.0006	0.0003
86	12	2856	TPVNSMGIN	0.0001				
86	12	2856	VPESDAA	0.0022				
86	12	1940	VPESDAA	0.0001	0.0001	0.0010	0.0001	0.0003
88	12	1940	WPILLILL	0.0021				
86	12	799	WPYRLMIV	0.0001				
100	14	616	76					

Table XII

HCY B27 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
AKHAMWFI	1767	8	12	86
AKNEVFCV	2593	8	12	86
ARALAHGV	148	8	14	100
DRSEL SPL	663	8	11	79
EKGAGPA	2603	8	11	79
EKMALYDV	2624	8	12	86
FKOKLGL	1733	8	12	86
GHRMAMDM	315	8	13	93
GKSTKPA	1240	8	12	86
GKPPARU	2806	8	11	79
HRMAYUDMM	316	8	13	93
KGSNLI	1390	8	11	78
LAIGVHT	1283	8	11	79
KKCELEAA	1403	8	14	100
KKKCELA	1402	8	14	100
U-KPTPL	1623	8	11	79
LHONNDV	697	8	12	86
LDLAVAV	988	8	11	79
NIIVSPHY	1832	8	12	86
PKGKQPI	58	8	13	93
PKGSRSY	100	8	11	79
PRRSFRL	112	8	12	86
RIADVPV	1140	8	11	79
RIIPVNSW	2854	8	12	86
RKLGVPRL	2943	8	12	86
RKPARUV	2607	8	11	79
RRCRAGV	2730	8	13	93
RKGTGLV	38	8	12	86
RRPQDMF	17	8	13	93
SKKCELE	1401	8	14	100
SNLGVKI	118	8	12	86
THDAHL	1571	8	13	93
TKLITPI	2985	8	12	86
TKVPAAYA	1243	8	14	100
TRCFDSV	2674	8	11	79
TRGVAKAV	1191	8	14	100
VRVCEKMA	2620	8	13	93
VRLEDSV	155	8	14	100
VRGLDVS	1423	8	11	79
ARHTPVNSW	2853	8	11	79
ARLVFPDL	2810	9	12	86
ARLVWLATA	1346	9	11	79
ARMMLTHF	2874	9	11	79
ARPDVNPPL	2298	9	11	79
DRSEL SPL	663	9	11	79

HCV B27 Super Motif

Sequence	Position	Pepitide No	No of Amino Acids	Sequence Frequency	Conservancy (%)
EKMALYDVV	2624		9	12	86
FKOKALGL	1733		9	12	86
GIRMAWDM	315		9	13	93
GKSTKVPAA	1240		9	12	86
GKRPALIV	2608		9	11	79
HRMAWDM	316		9	12	86
IKGRIILF	1390		9	11	79
KKKDELA	1402		9	14	100
UKGLSAFSL	2919		9	11	79
UIGTPLY	1623		9	11	79
UHSYSGEI	2927		9	11	79
IKSSSGRL	1166		9	12	86
UTKLGVPL	2942		9	12	86
NIWSPTHV	1932		9	12	86
NIRKQK	16		9	11	79
PRRPLGV	38		9	13	93
FTIPVNSWL	2854		9	12	86
FMWFGGA	1909		8	11	79
RKPAALIV	2607		9	12	86
FRGRASGL	2730		9	12	86
FRSHLGV	114		9	12	86
SKKCDLA	1401		9	14	100
TIIVPESDA	1937		9	12	86
TKVPAAYAA	1243		8	11	79
TRIADIVP	1139		9	11	79
THVESNKV	2251		9	12	86
WRTGGGI	22		9	13	93
VRVCEKML	2620		9	14	100
WRLLAPITA	1028		9	11	79
WRQMGON	2242		9	12	86
VRGLDVSI	1423		9	14	100
VRGRASGV	2729		9	13	93
ARALAHGVIV	148		10	14	100
ARAQAPPSW	1600		10	11	79
ARHTPVNSWL	2853		10	11	79
ARMILTHFF	2874		10	12	86
QSKKKDEL	1399		10	14	100
DRQSELSP	661		10	11	79
DRSELSP	663		10	11	79
EXGRAPAL	2603		10	11	79
FRVAICTRGV	1185		10	12	86
GIRMAWDM	315		10	12	86
GKSTKVPAA	1240		10	11	79
GKRPALIV	2606		10	11	79
KHAMWFSGL	1768		10	13	93

HCV B27 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
KKCDELAAKL	1403	10	12	86
LHONNDVOY	697	10	11	79
LKSSSGGRL	1168	10	12	86
OKALGILDTA	1735	10	12	86
RMGPGECAV	1909	10	12	86
RRGPRLGVNA	39	10	11	79
RRNGPGECA	1908	10	13	93
RRNSNLGKV	113	10	11	79
RRSNLGVNA	114	10	12	86
SKFGYGAQDV	2552	10	12	86
SKKKCDLAA	1401	10	14	100
THYVPESDAA	1937	10	14	100
TRGVAKAVDF	1191	10	12	86
TRVESBNKV	2251	10	11	79
VKFRGGQGV	22	10	12	86
VIVCEKMAKY	2620	10	13	93
VIMLEDGVNY	155	10	14	100
WRILAPITAY	1028	10	12	86
YKVLVLPVS	1254	10	11	79
YRRCRASGVL	2729	10	14	100
ALGVNALDGV	152	11	12	86
AQIMWNFTSGI	1767	11	13	93
ARALAHGVHVL	148	11	12	86
ARLIVFDLGV	2610	11	14	100
QISKKCODELA	1399	11	11	79
DFORSELSPL	661	11	14	100
EKGGRKPARU	2603	11	11	79
FRAAVCTRGVA	1185	11	11	79
GKSTIKVPAAYA	1240	11	11	79
GRVIDLTGCF	120	11	12	86
HRMAVDMMMNIV	316	11	12	86
KKKCODELAACL	1402	11	12	86
KKNTNRPODV	12	11	12	86
LHIGPTLYRL	1623	11	12	86
LHONNDVOYL	697	11	11	79
LKPTLHGPTPL	1619	11	11	79
LPRMNGPGECA	1907	11	11	79
PRGPRLGVNA	38	11	11	79
PRRNSNLGKV	112	11	13	93
RRNGPGECAV	1908	11	12	86
RRNSNLGKV	113	11	11	79
SRGNRNSPTIY	1929	11	12	86
SRNLGAVIDL	116	11	12	86
THYVPESDAAA	1937	11	12	86
VIMLEDGVNYA	155	11	12	86

HCV H27 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
YKVLNPSYA 136	1254	11	14	100

HCV B58 Super Motif

Table XIII

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
AALLRHIV	1904	8	13	93
AALAYCL	1673	8	12	86
AAQGYKL	1250	8	11	79
AATLGFCA	1264	8	14	100
AAVCTHGV	1187	8	12	88
ASLMAFTA	1793	8	11	79
ASSASQI	2204	8	14	100
ATLGFAY	1265	8	14	100
CSFSRL	172	8	12	86
CSGAYDI	1310	8	14	100
CSSNVSA	2819	8	11	79
CTQSSQL	1128	8	11	79
CTRGVAKA	1190	8	12	86
DTACGDI	994	8	12	86
DLTCGFA	124	8	11	79
EALENLV	750	8	11	79
EAMTHYSA	2794	8	14	100
ESDMAARV	1942	8	12	86
ETAGARLV	1342	8	12	88
ETTHRSPV	1207	8	13	93
FADLMGYI	130	8	14	100
FASRGNTV	1927	8	14	100
FSIFLLAL	174	8	11	79
FSYDIHCF	2670	8	11	79
FTEAMTRV	2792	8	14	100
FTSPVW	512	8	13	93
GAGVAGAL	1861	8	12	86
GATMGVLA	350	8	12	86
GALVGVW	1885	8	11	79
GARLVVLA	1345	8	12	86
GSGKSTV	1238	8	13	93
GSSQLYLV	1131	8	12	86
GSSCGATL	1188	8	11	79
GSSYGICV	2841	8	11	79
GTFPINAY	2083	8	11	79
HSYSPGEI	2928	8	11	79
HIPVNSML	2855	8	12	86
ISGIQYLA	1774	8	14	100
ITSCSNV	2816	8	14	100
ITWGADTA	989	8	12	86
KSTKVPAA	1241	8	12	86
LAGYGAGV	1857	8	11	79
LAGVIRVL	151	8	11	79
LAVAVEPV	972	8	14	100
LSAPSLKA	2211	8	11	79

HCV B58 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
LSPGALVV	1892	8	13	93
LSIGLHL	690	8	12	86
LTCGFADL	126	8	12	86
LTHDAHF	1570	8	13	93
MSADLEV	1654	8	11	79
NSWLGNI	2859	9	14	100
NTCVIOTV	1460	8	12	88
NINGSMH	416	8	13	93
PALSPGA	1889	8	13	93
PALSTGU	688	8	12	86
PLIWARMI	2870	8	11	79
PIPLLYRL	1628	8	14	100
QATVCARA	1595	8	13	93
RAHRTWFM	3019	8	14	100
RSELSPIL	664	8	11	79
RSNMGV	115	8	12	86
SAFSLSIS	2923	8	11	79
SSASOLSA	2206	8	14	100
STKVPAAV	1242	8	12	86
STLPGNPA	1784	8	14	100
STLPOAVM	2633	8	12	86
STYGFILA	1299	8	12	86
TAACGDII	995	8	12	86
TAGANLIV	1343	8	12	86
TTMRSPVF	1208	8	11	79
TTSCGNL	2738	8	12	86
VAGALVAF	1864	8	12	86
VTRHADVI	1138	8	11	79
VISTWLV	1681	8	12	86
WAKIMMNF	1766	8	14	100
WAKVLIVM	368	8	12	86
WAPGYPW	76	8	11	79
YMAOGYKV	1249	8	11	79
YSEPLDL	2905	8	12	86
YSTYCKRL	1298	8	11	79
YTNWDOOL	1106	8	16	114
AAKLDDCTM	2758	9	11	79
MAOGYKLV	1250	9	11	79
AAALAHGV	147	9	14	100
AAILGFAY	1264	9	11	79
AAVCTHGA	1187	9	13	83
ASOLSAFSL	2208	9	26	186
ATLFGAYM	1265	9	11	79
ATVCARAQA	1596	9	13	93
CAALLRIV	1903	9	13	93

HCV H58 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
CAYVELTPA	1530	9	11	79
CSFSRLA	172	9	14	100
CSGGAYDIH	1310	9	12	86
CTGSSSLY	1128	9	11	79
CIHQVAKAV	1190	9	11	79
CTWNNSTGF	555	5	11	79
DAGCAYEL	1527	9	11	86
DTAACGDII	994	9	12	93
DIHCFDSTV	2673	9	13	86
ETAGARLV	1342	9	12	86
ETMNSPVF	1207	9	12	86
FSRLALL	174	9	14	100
FSIDPTFI	1469	9	14	100
FTGLTHDA	1567	9	13	93
GAGVAGALV	1861	9	12	88
GALVAFKIM	1866	9	12	86
GALVAFKVM	1866	9	14	100
GAYQMMWRL	1916	9	14	100
HSKKKDEL	1400	9	14	100
HTPGCVPCV	222	9	11	79
ITWGAADTAA	989	9	12	86
ITYSYSGKF	1296	9	12	86
KAGLLDTA	1736	9	12	86
KSTKVPAAY	1241	9	12	86
LAALAAYCL	1672	9	12	86
LAQFKOKA	1729	9	14	100
LAGLAYYSM	356	9	11	79
LAQYGAGVA	1857	9	11	79
LSAFLSHSY	2922	9	14	100
LSTLPGNPA	1783	9	24	171
LTCGFADLM	126	9	14	100
LTDPSHITA	2180	9	12	86
LIGRDKNOV	1052	9	13	93
LTHDAHFL	1570	9	11	79
LTTSCGNIL	2738	9	12	86
MAKNEFCV	2592	9	12	86
MAWDMAMNV	318	9	13	93
NAVAVYRGL	1418	9	14	100
NSLLRHHNM	2481	9	14	100
NSWLGNIIM	2858	9	24	171
NINRRPODV	14	9	12	86
PALSPGAL	1889	9	13	93
PSVAATIGF	1261	9	14	100
PTLKGPTPL	1621	9	11	79
PTLWARMIL	2870	9	11	79

HCV B58 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
QAEIAGARL	1340	9	12	86
RAVCTRGV	1186	9	12	86
RAIAGIVR	148	9	14	100
RAQAPPSW	1601	9	11	79
RAYAMOREM	811	9	16	114
RSEI SPILL	664	9	11	79
RSRLGKM	115	9	12	86
SSSASQLSA	2205	9	14	100
STKVPAAVA	1242	9	12	86
STLPGNPAI	1784	9	11	79
STWLVGCV	1663	9	12	86
TAGARLVVL	1343	9	14	100
TSCSSNWSV	2817	9	14	79
TTMAKNEV	2589	9	11	100
VAAITLFGA	1263	9	14	100
VAGGIVVQM	933	9	14	100
VAYDITVCA	1592	9	12	86
VAYYHGLDV	1420	9	14	100
VSILPOAYM	2832	9	12	86
VTOIVDFSL	1463	9	12	86
WAKIMMNI	1766	9	11	79
YAAQGYKVL	1249	9	11	100
YAPILWARM	2668	9	14	79
YSPGENRIV	2930	9	11	79
YSPQINIEF	2848	9	11	86
YSTYGKFLA	1298	9	12	86
YTNVDQILV	1106	9	11	79
YAGGYKVL	1250	10	11	186
YAILGFGAYM	1264	10	28	86
ASLHVITEAM	2787	10	12	100
ASSSASQLSA	2204	10	14	93
ATGNLPGCSF	165	10	13	100
CSFSILAL	172	10	14	79
CTGSSQLYL	1128	10	11	79
DARVCAQLWM	733	10	18	129
DSVIDNTCV	1454	10	12	86
DTLTGFAEL	124	10	12	86
EMALWROEM	2237	10	24	171
ETAGARLVVL	1342	10	12	86
FADLMGYPL	130	10	11	79
FTEAMTRYSA	2792	10	14	100
GAARALAHGV	146	10	11	79
GADTAACGDI	992	10	12	86
GAGVAGALVA	1861	10	12	86
GALWGWVCA	1895	10	11	79

HCV B58 Super Motif:

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
GARLVVLAIA	1345	10	11	79
GAYQWNNRIJ	1916	10	14	100
GGGKSTKVPA	1238	10	12	88
GTILDOAETA	1335	10	14	100
HSKKKCDJEA	1400	10	14	100
IAFASRGHIV	1925	10	14	100
ISGIQYLAGL	1774	10	14	86
ITRVESENKV	2250	10	14	100
ITSCSNVSV	2818	10	11	79
ITYSTYGFEL	1296	10	12	86
KSTKVPAAYA	1241	10	11	79
LADGGCSGGA	1305	10	12	88
LAEDFKQKAL	1729	10	12	86
LAI PPRAYAM	806	10	13	93
LSFGALVGV	1892	10	11	79
LSPRGSPSW	88	10	11	100
LSRAIRPIWFM	3017	10	14	79
LSILPGNPAI	1783	10	11	79
LTHPIIKYIM	1642	10	16	114
NICVOTIOTDF	1460	10	12	86
PAILSPGALV	1889	10	12	86
PALSTGLHL	688	10	12	86
PARLIVPPDL	2609	10	11	79
PSWQDMWKCL	1607	10	11	79
PTGSGKSTIV	1236	10	13	93
PTIIVYPESDA	1936	10	12	86
PTUIGPTPL	1621	10	11	79
PTLWARMLLM	2870	10	22	157
PTLLYHLGA	1628	10	13	93
QAEIAGARLV	1340	10	12	86
QAPPSWQDM	1603	10	24	171
OATVCAIAGA	1595	10	11	79
RAAKLODCTM	2757	10	16	114
RAAVCTRGVA	1186	10	11	79
RALAHGVRL	149	10	11	100
SASQISAPSL	2207	10	14	93
STKVPAAAYAA	1242	10	13	86
STIWLVGVL	1663	10	12	86
TAGARLVVLA	1343	10	12	79
TARHTPVNSW	2852	10	11	100
TSCSSNSVYA	2817	10	14	93
TSMITDPSHI	2177	10	13	86
TSMWLVGV	1662	10	12	79
TTIMAKNEIVF	2589	10	11	79
TTLPALSTGL	685	10	11	79

HCV B58 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
VAATLGFAY	1263	10	14	100
VTGETPSGM	1507	10	16	114
VTRIADVPV	1138	10	11	79
WAQPGYPMPL	76	10	12	86
WATMLMLTHF	2873	10	11	79
WAPDYNPL	2297	10	11	79
YAAQGYKVLV	1249	10	11	79
YSPGEINRA	2930	10	11	79
YSPQRMEL	2648	11	11	79
AAALALIGVIV	147	11	12	86
AASLIVTEAM	2788	11	11	79
AAVCTRGVAKA	1187	11	14	100
ASHLPYIEGGM	1717	11	11	79
ASQLSAPSLKA	2208	11	11	79
CARQAQPPPSW	1598	11	14	100
CSIFILAIL	172	11	11	79
CTGSSDLVIV	1128	11	11	79
CTRGVAKAVDF	1190	11	16	114
DARVCAQLWMM	733	11	24	171
DTLTGCFADLM	124	11	12	86
ETAGARLVLA	1342	11	11	79
FADLMGYPLV	130	11	11	79
FSLHSTSPGEI	2925	11	13	93
FTGLTHIDAHF	1567	11	79	79
FTTLPALSTGL	884	11	12	86
GADTAACDII	992	11	12	86
GAGVAGALVAF	1861	11	11	79
GALVVGWCA	1885	11	14	100
GAYQMMNIIA	1818	11	12	86
GSCKSTKVPAA	1238	11	14	100
HSKKKCEIAA	1400	11	11	79
HSYSPGEINRV	2928	11	11	86
HTPNSMIGNI	2855	11	12	86
IIIVSEENKV	2250	11	14	100
ITSQSSNNSVA	2816	11	11	79
IITYSTYGKFLA	1296	11	11	79
KSTKVPAAAYAA	1241	11	11	79
LADGGSGGAY	1305	11	11	79
LAGYGAGVAGA	1857	11	14	100
LSNSILPPIIM	2479	11	11	79
LSPGALVGV	1892	11	11	79
LTCGFADLMGY	126	11	12	86
LTSALTPSIH	2176	11	13	93
NAVAVRGIDV	1418	11	13	93
NTRRPDOYVF	14	11	11	79

HCV B58 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
PALSPGALVY	1889	11	12	86
PSVAATLGFGA	1261	11	14	100
PTDPRARRNL	109	11	12	86
PTIIVPESDMA	1936	11	12	86
PTLHGPTFLY	1821	11	11	79
PTPLYLGLAV	1628	11	13	93
QAETAGARLV	1340	11	12	86
QAPPSWDMMV	1603	11	11	79
QIVDFSLDPIF	1465	11	12	86
RSQTHGRROA	55	11	13	93
RSQTHGRROA	55	11	11	79
SADLEVTSIV	1655	11	11	93
SSASQLSAPSL	2206	11	13	86
SSDIYLVTRHA	1132	11	12	86
STWLVGGVLA	1663	11	12	79
TARIIPVNSWL	2852	11	11	86
TSLTGDKNOV	1050	11	12	86
ISTWLVGGVL	1662	11	11	79
TLPLALSTGL	685	11	12	106
VAAATGFGAYM	1263	11	26	100
VAGALVAFKVM	1864	11	14	86
VAVEPVFSDM	974	11	12	79
VAYQATVCARA	1592	11	11	100
VAYYRGLDVSV	1420	11	14	86
VTSIWLVGCV	1661	11	12	86
WAQPGYPMFLY	76	11	12	86
WARMLMTHFF	2873	11	12	86
YAAQGYKVL	1249	11	11	79
YATGNI PGCSF	164	11	12	86
YTNMDQLVGM	1106	11	11	79

289

HCY B62 Super Motif

Table XIV

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
ALLSPGAL	1880	8	13	93
ALAKGVRV	150	8	14	100
ALGILQTA	1737	8	12	88
APLWARM	2869	8	11	79
AQAPPSW	1602	8	12	86
AQGYKLV	1251	8	11	79
AVAYVRGL	1419	8	14	100
AVCTIRGVA	1188	8	11	79
AVQWNRRL	1917	8	14	100
CLWMMILL	739	8	12	86
QMSADLEV	1853	8	11	79
QDHLIFW	1556	8	12	86
CVIOTVDF	1462	8	12	86
DILAGYGA	1855	8	12	86
DLOGSYFL	279	8	11	79
DLMGVPL	132	8	11	79
DLVNLIPA	1883	8	12	86
DOAELAGA	1338	8	13	93
EIPFYGA	1377	8	12	86
EGRKOKAL	1731	8	12	86
EVVSTIW	1659	8	14	100
FISQIQY	1773	8	11	79
FDLGVRV	2615	8	14	100
FRGGGV	24	8	12	86
FOVNLIIA	1228	8	14	100
GKYLGL	1776	8	11	79
GLDLAVA	968	8	13	93
GRPLGVRA	41	8	14	100
GQWGGVY	28	8	12	86
GVAGALVA	1863	8	11	79
GVAKAVDF	1193	8	11	79
GVLAALAA	1670	8	12	86
GVAVCEKM	2519	8	14	100
GVVCAAIL	1900	8	11	79
HVQGECA	1910	8	12	86
HVSPTIIV	1933	8	12	86
ILGGVAA	1816	8	12	86
ILGIGTVL	1331	8	13	93
ILSPGALV	1891	8	12	86
IMAKNEVF	2591	8	13	93
IPIFYKAI	1378	8	11	79
IPLVGARL	137	8	11	79
INDVQVLY	701	8	12	86
INPPDLGV	2613	8	11	79
IVGGVILL	30	8	13	93

ICV B62 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
KMALYDVV	2625	8	12	86
KPARLIVE	2608	8	12	86
KOKALGL	1734	8	12	86
KVPAYVAA	1244	8	11	79
LLEANI LW	2235	8	12	86
LINTNGSW	414	8	11	79
LLALISCL	178	8	12	86
LLAPITAY	1030	8	14	100
LLADARV	729	8	13	93
LLYRLGAV	1629	8	13	93
LMGYPLV	133	8	11	78
LPAISTGL	687	8	14	100
LPGCSFSI	168	8	13	93
LPRRGPH	37	8	13	93
LPRVODHL	1553	8	13	93
LPVEOGM	1720	8	12	86
LQDCTMLV	2761	8	12	86
LVAYOATV	1591	8	12	86
LVDILAGY	1853	8	11	79
LVGGLAA	1667	8	12	86
LVLNPSVA	1257	8	14	100
LVLNLPAL	1884	8	11	78
LVTIHADV	1137	8	12	86
LWGVVCA	1897	8	11	79
LWVCESA	2773	8	11	79
MLMTIFF	2878	8	12	86
MLTDPSH	2179	8	14	100
NILGMYVA	1815	8	12	86
NIVDOYTL	700	8	12	86
NILWROEM	2239	8	12	86
NPSVAATL	1260	8	14	100
PLGGAARA	143	8	11	79
PLTYRLGA	1628	8	13	93
PPPSWIDOM	1605	8	12	86
PPSWIDOM	1606	8	11	79
PMHIGPCL	2318	8	11	79
QWGGVTL	29	8	13	93
QALRIPOA	336	8	12	86
QPEYDLE	2808	8	12	86
QPGYPMFL	78	8	11	79
RUHQLSAF	2918	8	12	86
RLVFPOL	2611	8	11	79
RLAPITA	1029	8	12	86
RLVVLATA	1347	8	12	86
RMAWDMMM	317	8	12	86

HCY B62 Super Motif

Sequence	Position	No of Amino Acids	Sequence Frequency	Conservancy (%)
RMILMTHF	2875	8	12	86
RPDYNPPL	2299	8	11	79
ROBAGCN	2243	8	12	86
RVCCKMAL	2521	8	14	100
RVESEKIV	2292	8	12	86
RWGDRIWV	2100	8	11	79
SIFLLALL	175	8	14	100
SLDPTFTI	1470	8	14	100
SPGENTIV	2931	8	11	79
SPGQNEF	2649	8	11	79
SQLSAPSL	2209	8	13	93
SVAATLGF	1262	8	14	100
TIMAKNEV	2590	8	11	79
TLGFQAYM	1266	8	13	93
TLHGPTPL	1622	8	11	79
TLRGNPAI	1785	8	11	79
TLWARHML	2871	8	12	86
TPCSGSMI	1875	8	12	86
TPGCYPCV	223	8	12	86
TOIYDFSI	1464	8	11	79
TYCARADA	1597	8	11	79
VIDQNTCV	1456	8	12	86
VLAALAAY	1671	8	12	86
VACECYDA	1521	8	13	93
VLDQAETA	1337	8	14	100
VLEDGANT	157	8	14	100
VLINPSVAA	1258	8	12	86
VLVQGVLA	1666	8	14	100
VLVNPSV	1258	8	14	100
VMGSSYGF	2639	8	11	79
VPIESDAAA	1940	8	12	86
VQMMNRIU	1918	8	14	100
VVAIDALM	1439	8	11	79
VVGWCAAA	1898	8	11	79
VVTSTWVL	1660	8	12	86
WMNRILAF	1920	8	14	100
WPLILLLL	799	8	12	86
WMLVGGVL	1665	8	14	100
YLAGLSTL	1779	8	14	100
YPRILMIV	616	8	12	86
YPIESDAA	1938	8	12	86
ALLSPGALV	1890	9	12	100
ALAHQVRVL	150	9	14	100
ALSTGLHL	688	9	12	86
ALVWGVC	1896	9	11	79

HCY R62 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
APPSWDQM	1604	9	12	86
APTLWARM	2869	9	11	79
AQGYKLV	1251	9	11	79
AQGYPMFL	77	9	12	86
AVQWNRLL	1917	9	14	100
CMASDLEW	1653	9	11	79
ICGSRFLV	279	9	12	86
DEEVTSTW	1657	9	11	79
DLMGYRLV	132	9	11	79
DVNLPLAI	1883	9	11	78
DLVWICSA	2772	9	12	86
DLVTRHAA	1134	9	11	79
DPQLSDXSV	2410	9	12	86
DPRRSRNL	111	9	13	93
EPFYGKAI	1377	9	12	86
EMGNITRV	2245	9	12	86
EVVSTWVL	1658	9	14	100
FSGIQYLA	1773	9	12	86
FLALLSCL	177	9	13	93
FLLDADRV	728	9	11	79
FOYRQGRV	2646	9	14	100
GIGTMDQA	1333	9	13	93
GLPVQDHL	1552	9	11	79
GLHDLAVAV	968	9	13	93
GLTHDAHF	1569	9	12	86
GRCEGAVW	1912	9	14	100
GPTPLLYRL	1625	9	13	93
GQVGGVTL	28	9	12	86
GVAGALVAF	1863	9	12	86
GVLAAALAAV	1670	9	11	79
GVNVAIGNL	161	9	14	100
GVNVECKMA	2619	9	13	93
GVNLEDGV	154	9	12	86
HLHQNINOV	696	9	11	79
HLPYIEQAM	1718	9	13	93
HMNNFISGI	1769	9	11	79
HQNNINOV	698	9	11	79
HWGRGEGAV	1910	9	11	79
ILAGYGAGV	1856	9	13	93
ILSPGALVV	1891	9	14	100
KVLVLPSPV	1255	9	14	100
LITSQSSNV	2815	9	11	79
LVPDLGV	2612	9	14	100
LLFLLADA	726	9	12	86
LLFNILGAV	1812	9	12	86

HCV B62 Super Motif (No binding data)

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
LLPRIGPL	36	9	13	83
LPALSPGA	1888	8	13	93
LPALSTGL	687	9	12	86
LPCEPEPV	2165	9	12	86
LPCCSFIF	169	9	13	93
LVGWLAL	1667	9	12	86
LVLPVSAA	1257	9	14	100
LVNLLPAL	1884	9	11	79
LVTRHADV	1137	9	11	79
LVGVVCAA	1897	9	11	79
NILGWMVA	1815	9	12	86
NIRIGVRI	1282	9	11	86
NMDVOYLY	700	9	12	86
NIGKVIDTL	118	9	12	86
NIPGCSFSI	168	9	13	93
NDOOLVGN	1108	9	11	79
PIGGAMIAL	143	9	11	79
PILYHIGAV	1628	9	13	93
PPPSMDQMW	1605	9	11	79
PPWIGGRL	2317	9	11	78
POPEYDIL	2807	9	11	86
PWQDRIEF	1554	9	12	86
PVNSMIGNI	2857	9	14	100
QWGVNLL	29	9	13	93
QLSAPSLKA	2210	9	13	79
QPEYDEI	2808	9	11	79
QRYPMFLY	78	9	12	86
QRTGRQI	57	9	13	93
RLAPITAY	1028	9	12	86
RLMLMTHFF	2875	9	12	06
RVCCKMAY	2621	9	14	100
RVESENKV	2252	9	12	86
RMLEDGNY	156	9	12	86
SMALTOPSH	2178	9	14	100
SPGALVGV	1893	9	13	93
SPGEINRYA	2931	9	11	79
SPQQRVER	2649	9	11	79
SPRGSRPSW	99	9	11	79
SVDCNTCV	1455	9	12	86
TIMAKNEVF	2590	9	11	79
TUHGPIPL	1822	9	11	79
TLPALSTGL	686	9	11	79
TLTCGFADL	125	9	12	86
TLVAMMLM	2871	9	11	79
TLLYHLGA	1627	9	13	93

HCV B62 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
TALDOAETA	1336	9	14	100
VIDITICGF	122	9	12	86
VLEDGWNVA	157	9	12	86
VLVDLAGY	1852	9	11	79
VLVGVLA	1668	9	12	86
VLINPSVA	1258	9	14	100
VOMANRLA	1918	9	14	100
VGVVCAAI	1898	9	11	79
VVTSTWLV	1660	9	12	86
WMNRLAFA	1920	9	14	100
WLVGVLA	1665	8	12	86
YPLVGAPL	136	9	11	78
YLVAQAIV	1590	9	12	86
YLVTIRIADV	1136	9	12	86
YDATTGAAH	1594	9	13	93
YGDGGSV	276	8	12	86
YGVGBFLL	637	9	13	93
YVPESDAA	1939	8	12	86
ALSPGALVV	1880	10	12	79
ALVGVVCAA	1896	10	11	79
APPSWDDMW	1604	10	11	79
APILWMARML	2869	10	12	86
ADPGVPMILY	77	10	14	100
AVAYRIGLDV	1419	10	11	79
AVCTRGVAKA	1188	10	11	79
AVOMANRLA	1817	10	14	100
CLHKLGVPL	2941	10	12	86
CVTQIVDFSL	1462	10	12	86
DILAGYGAGV	1855	10	11	79
DLEWISTWV	1657	10	12	86
DLCVRNCEAA	2617	10	13	93
DLSDGSWSIV	2412	10	11	79
DLVNLIPAIL	1803	10	11	79
DOAETAGARL	1339	10	12	86
DWPFPGGGQI	21	10	12	100
ELTSCSNV	2814	10	14	100
EOHKQKALGL	1731	10	12	86
EVVSTWLV	1659	10	12	86
GLSAFSLHSY	2921	10	11	79
GLSTLPGNPA	1782	10	14	100
GLTHIDAFIL	1569	10	13	93
GRGEGANQMA	1912	10	12	86
GOVGGVYLL	28	10	13	93
GVOMTVYHGA	1081	10	11	79
GVNCEKEMAL	2619	10	14	100
	10504			

HCY B62 Super Motif

Sequence	Position	Pepide No.	No. of Amino Acids	Sequence Frequency	Conservancy (%)
IKNIWDVOYL	698		10	11	79
ILAGYGAGVA	1856		10	11	79
ILGGWVAACL	1816		10	12	86
IMANVEFCV	2591		10	11	79
KOYLAGESTL	1777		10	14	100
INFPDLGVRV	2813		10	11	79
KPTLHGPTPL	1620		10	12	86
KVIDLTGCF	121		10	14	100
KVLVLRPSVA	1255		10	12	86
LLFNILGGWV	1812		10	13	93
LLPAILSPGA	1887		10	11	79
LMGYPLVGA	133		10	13	93
LPAILSPGAL	1888		10	13	93
LPGCSFSL	169		10	13	93
LPRHGHILGV	37		10	12	86
LPVODHLEF	1553		10	12	86
LVAYQATVCA	1591		10	11	79
LVDLIAGYGA	1853		10	12	86
LVGGVLAALA	1667		10	11	79
LVGGVCAAI	1897		10	14	100
MLTDPSHITA	2178		10	14	100
MPGCSFSF	168		10	13	93
NPVVAATLGF	1260		10	14	100
PITYSTYGEF	1295		10	11	79
PLGGAARALA	143		10	11	79
POREYDAEI	2807		10	12	86
PVODHLEW	1554		10	14	100
PVNSMLGNIL	2857		10	13	93
PVYCFPSPV	508		10	12	88
QAPCEPDDV	2164		10	11	79
QPEKGGAKPA	2601		10	11	79
RIIGLSAFSL	2918		10	11	79
RLNEPDLGV	2611		10	12	86
RMAYMDMMNNW	317		10	12	86
RNLEDGWNIA	156		10	11	79
SUSYSRGEI	2926		10	12	86
SLTGDKKQCV	1051		10	11	78
SPGALVVGIV	1893		10	11	79
SQLSAPSLKA	2208		10	13	93
SQPRGRHQR	56		10	14	100
SVAAITLFGA	1262		10	11	79
TUHGPTPLY	1622		10	12	86
TILFNILGGW	1811		10	11	79
TILPALSTGL	686		10	11	79
TILCGFADLM	125		10	12	86

HCY B62 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
TPCTGSSSL	1126	10	11	79
TPPLYRLGAV	1627	10	13	93
TPVNSMIGNI	2856	10	12	86
TVDFSLDPTF	1466	10	12	86
VIDTITCGFA	122	10	12	86
VLAALAYCL	18/1	10	12	86
VLDQETAGA	1337	10	14	100
VLNPSVAATL	1258	10	11	79
VLTTSCGNTL	2737	10	12	86
VLVGVLAL	1666	10	14	100
VLVNPVSAA	1256	10	11	86
VAKSSYGFQY	2639	10	12	79
VPESDAARV	1940	10	14	100
VQWNNRLAF	1818	10	11	79
VGVVCAIL	1898	10	12	86
VWLVGVILAA	1665	10	12	86
VXGSSGRL	1165	10	13	93
VILPRGRRL	35	10	11	79
VLVTRHADV	1136	10	12	86
VWGLDSSF	276	10	11	79
ALVGVVCAIL	1896	11	13	93
APTGSCKSTK	1235	11	11	79
APTLWARMLM	2889	11	12	86
ADAPPSSMDM	1602	11	11	79
AVCTRGVAKAV	1188	11	14	100
AVQWNNRLAF	1917	11	11	79
DLAGYGAGVA	1855	11	12	86
DLFWTSTWIL	1657	11	13	93
DLGVRVCEXMA	2617	11	11	79
DLMGYRLVGA	132	11	12	86
DLVLTTHADV	1134	11	12	86
DOAETAGATLV	1339	11	12	86
DWTFPGGCON	21	11	12	86
EOROKALGIL	1731	11	14	100
FISGIYLAGL	1773	11	11	79
FLADGCGSCGA	1304	11	14	100
FRGGQDVGIV	24	11	11	79
KQYSGQRNRF	2646	11	14	100
GKQYLAGLSTL	1776	11	12	86
GLPYQDHLRF	1552	11	11	79
GLSTPGNPNL	1782	11	13	93
GPTPLYRLGA	1625	11	13	93
GPVYCFIPSPV	507	11	12	86
GVLAALAYCL	1670	11	14	100
GVRVCEXMAV	2619	11	14	100

HCY B62 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
GVNLEDGVNY	154	11	12	86
H1HONVDVQY	696	11	11	79
1MMNFISSQIY	1769	11	13	93
HONVDVQY	698	11	11	79
HNGFGGAGQW	1910	11	11	86
ILGGVVAQOLA	1610	11	12	95
ILGIGTMDQA	1331	11	13	93
ILSPGALWGV	1891	11	11	79
KPAHLVFPDL	2608	11	11	79
KPTLHGPTPL	1620	11	12	86
KOKALGLOTA	1734	11	12	86
KVIDLTGFA	121	11	14	100
KVLVLPNSVAA	1255	11	14	100
LVAFASRGNIY	1924	11	14	100
LITSCSSWSY	2815	11	11	79
LIVFPOLGVNY	2812	11	13	93
LIPLILADAHV	726	11	12	86
LIENILGGWA	1812	11	13	93
LIPLSPGAL	1887	11	13	83
LIPIRGPTLGV	36	11	11	79
LLSPRGSPSW	87	11	11	86
LLVROGNGGN	2240	11	12	86
LPAILSPGALV	1888	11	12	86
LPALSTGLIHL	687	11	13	93
LPQCSFSIHL	168	11	12	86
LPVQDQLEPW	1553	11	12	86
LVGGVLAALAA	1667	11	14	100
LVLPNSVAATL	1257	11	11	79
LVTRHADVPV	1137	11	11	88
LVGGVCAAIL	1897	11	12	86
LVGGVCAAIL	1815	11	12	86
NITFESQKV	2249	11	13	93
NILPGWVAQOL	1888	11	13	93
NILPAILSPGA	168	11	11	79
NLPQCSFSIHL	1295	11	13	93
PIVSTYSGRL	2403	11	11	79
PLGEGKDTOL	2667	11	11	78
PMGFSYDITRCF	1606	11	11	86
PPSWDQAMKQL	2857	11	12	86
PVNSWLGNILM	508	11	13	93
PVYCTPSPV	635	11	12	86
PLMWGVEHIL	2243	11	12	86
POEMGANTIV	2821	11	12	86
PVCEKMLYDV	175	11	14	100
SIFLALISQL	2178	11	14	100
SMLTDPSHITA				

HCY B62 Super Motif

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)
SPITIVPESDA	1835	11	12	86
SQLPCEPEPV	2163	11	12	86
SVAATIGFGAY	1262	11	14	100
TLGFGAYMSKA	1266	11	12	86
TLILNLGGWV	1811	11	12	86
TPCTGSSDL Y	1126	11	11	79
TPQLPVCOX L	1550	11	13	93
TPVNSWLGNI L	2856	11	12	88
TVLDOAETAGA	1336	11	12	86
VLCECYDAACA	1521	11	11	79
VLVDLAGYGA	1852	11	11	86
VLVGGVLAALA	1666	11	12	79
VOPKGGRRPA	2600	11	11	100
VQWNNRILAF A	1918	11	14	78
VVCALRRHV	1801	11	11	86
WMLVGGVLAAL	1665	11	12	86
YKSSSGRPL	1165	11	12	86
YLAVYQATVCA	1590	11	12	79
YQATVCARAOA	1594	11	11	88
YNGKOSVTL	278	11	12	88
YVPESDAAHV	1939	11	12	86

426

Table XV

HCV A01 Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0101
ASFCGSPY	166	26 0026	20	100	
DNSVLSIRKY	737	20 0255	18	90	0.0001
FAAPFTOCGY	631	20 0254	19	95	0.0680
GFAAPFTOCGY	630		19	95	
GRETMEY	140		15	75	
GYSLNFMGY	579	2 0058	17	85	
HILWKAGLY	149	1069.04	20	100	0.1100
KQAFETSPY	653	20 0256	19	95	0.0001
LLDTASALY	30	1069.01	17	85	12.0000
LSLDVSAIFY	415	1090.07	19	95	0.0150
LTFGRETMEY	137		15	75	
MMWVWGPSLY	360	1039.01	17	05	0.0810
MSITDLEAY	103	2.0126	15	75	0.8500
NSVLSIRKY	738	2.0123	10	90	0.0005
PLDKGKPY	124	1147.12	20	100	
PLDKGKPY	124	1069.03	20	100	0.1700
PTTGRTSLY	797	1090.09	17	85	0.2100
SASFCGSPY	165		20	100	
SILDVSAIFY	416	1069.02	19	95	5.2000
STTDLEAY	104		15	75	
TTGRTSLY	798	26 0030	17	85	
WLSILDVSAIFY	414	26 0551	19	95	
WMMWVWGSPS	359	1039.06	17	85	0.3200
YPAIMPLY	640	19 0014	19	95	
YSUNFMGY	580	26 0032	17	85	

25

Table XVI

ICV A03 Motif with Binding Information

Sequence	Position	No of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
AACNWTNGER	647	10	12	86	0.0003
AARALAHGVR	147	10	11	79	
AATLGFGA	1264	0	14	100	
AATLGFAY	1264	9	14	100	
AAVCTHGVAK	1187	9	11	79	
AAVCTHGVAK	1107	10	11	79	
AAVCTHGVAKA	1107	11	11	79	
ACNMTFRGER	648	9	12	86	
ADGGGCGGGA	1306	9	11	79	
ADGGGCGGAY	1306	10	11	79	
ADVPVRII	1142	0	12	06	
ADVPVRII	1142	9	11	79	
AFASRGNH	1926	0	14	100	
AGALVAFK	1065	0	12	06	
AGAHILVILA	1344	9	12	06	
AGARILVILATA	1344	11	11	79	
AGLSTLPGNPA	1781	11	14	100	
AGVAGALVA	1062	9	12	06	
AGVAGALVAF	1062	10	12	86	
AGVAGALVAFK	1062	11	12	86	
AGWLTSPH	94	0	12	86	
AGWLTSPHSR	94	11	12	06	
AGYGAGVA	1050	0	12	06	
AGYGAGVAGA	1050	10	12	06	
ALGLLOTA	1737	0	12	06	
ALSTGLIH	609	0	12	06	
ALSTGLIHII	609	10	12	06	0.0003
ALVGVVCA	1896	9	11	79	
ALVGVVCAA	1896	10	11	79	
ASLMAFTA	1793	8	11	79	
ASQLSAPSLK	2208	10	11	79	
ASQLSAPSLKA	2208	11	11	79	
ASRGNINSPTH	1928	11	12	06	
ASSASQLSA	2204	10	14	100	
ATGNLPCCSF	165	10	13	93	
ATLGFAY	1265	8	14	100	
ATLGFAYMSK	1265	11	12	86	
ATPKTSEI	48	0	11	79	
ATVCARAQA	1596	9	11	79	
AVCTHGVAK	1108	8	11	79	
AVCTHGVAK	1188	9	11	79	0.0260
AVCTHGVAKA	1188	10	11	79	
AVQWMNRLIA	1917	10	14	100	
AVQWMNRLIAF	1917	11	14	100	
CAALLRHII	1903	8	13	93	

HCV A93 Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
CAYELTPA	1530	9	11	79	
CGFAQMGY	120	9	13	93	
CGNLTICV	2742	8	11	79	
CGSSQLYLVR	1130	11	11	79	
CGYRRCRA	2727	8	14	100	
CIIRKGVPLR	2941	11	12	86	
CSFSIFLA	172	9	14	100	
CSSNVSA	2819	8	14	100	
CSSNVSAH	2819	9	12	86	
CTCGSSQLY	1128	9	11	79	0.0001
CTRGVAKA	1190	8	11	79	
CTRGVAKAVDF	1190	11	11	79	
CTWMSITGF	555	9	11	79	0.7600
CTWMSITGFK	555	11	11	79	0.0008
CVQPKGGR	2539	9	11	79	0.0011
CVQPKGGRK	2539	10	11	79	
CVITQIVDF	1462	8	12	86	
DAIFLSQIK	1574	9	14	100	0.0003
DALVWICESA	2771	10	11	79	
DFSLDPTF	1468	8	14	100	
DGQSSQA	1307	9	11	79	
DGQSSQAY	1307	9	11	79	
DGQSSQAY	1316	9	12	86	
DIIICDECH	1855	8	12	86	
DILAGYA	1855	11	11	79	
DILAGYGAGVA	1855	9	13	93	0.0003
DILAGYCEK	2617	11	13	93	
DILGVRCEKMA	2617	11	11	79	
DLMGPIPLVGA	132	8	11	79	
DLVNLTPA	1803	9	11	79	
DLVNLTPA	2772	9	11	79	
DLVWICESA	1134	8	12	86	
DLVLTIRH	1134	9	12	86	0.0003
DLYLVTRHIA	124	8	12	86	
DILTCGFA	1143	8	11	79	
DVIPRRR	1143	8	14	100	
EMATRYSA	2794	8	14	100	
ECYDAGCA	1524	8	11	79	
ECYDAGCAWY	1524	10	11	79	
EDLVNLTPA	1882	9	11	79	
EGAVQMMAN	1915	9	14	100	0.0004
ELPFYGA	1377	8	13	93	
EMGNIR	2245	8	12	86	
ETAGARLVILA	1342	11	12	86	
ETMRSPVF	1207	9	12	86	0.0008
EVFOKPEK	2596	9	12	86	
FOVQPKGGR	2598	10	11	79	

HCV A03 Motif with Binding Information

Sequence	Position	No of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
HCYKPKGGRK	2590	11	11	79	
FGAYMSKA	1269	8	12	86	
FGAYMSKAH	1269	9	12	86	
FGCTWNSIGF	553	11	11	79	
FGYGMQDVR	2554	9	12	86	0.0008
FIGIQYLA	1773	9	14	100	
HLAKKCGSGA	1304	11	11	79	
FILLADAR	728	8	14	100	
FSYDIRCF	2670	8	11	79	
FTEAMTRY	2792	8	14	100	
FTEAMTRYSA	2792	10	14	100	
FTGLTHIDA	1567	9	13	93	
FTGLTHIDAI	1567	10	13	93	
FTGLTHIDVIF	1567	11	13	93	
GAARHAIH	146	0	11	79	
GAARHAIHVR	146	11	11	79	
GAGVAGALVA	1061	10	12	86	
GAGVAGALVAF	1061	11	12	86	
GATWGVLA	350	8	12	86	
GALVGVCA	1895	10	11	79	
GALVGVVCA	1895	11	11	79	
GALVGVCA	1345	8	12	86	
GARLVLA	1345	10	11	79	
GARLVVLA	1345	11	11	79	
GAQNMNI	1916	8	14	100	
GAQNMNIILA	1916	11	14	100	
GAVMKAI	1270	0	12	86	
GCAYVELTPA	1529	10	11	79	
GCSFSIFLA	171	10	14	100	
GCTWNSIGF	554	10	11	79	
GDDLVICSA	2770	11	11	79	
GILCGSVF	278	8	12	86	
GFDLMGY	129	8	13	93	
GFGAYMSK	1268	8	12	86	
GFGAYMSKA	1268	9	12	86	
GFGAYMSKAH	1268	10	12	86	
GFOYSGQR	2645	9	11	79	
GISYDIRCF	2669	9	11	79	
GGAATIA	145	8	11	79	
GGAATIAIH	145	9	11	79	
GQCSQAY	1308	8	11	79	
GQGMGGAY	26	10	14	100	
GQHVQMA	935	8	11	79	
GQNGGVY	27	9	14	100	
GRIIFCI	1392	9	14	100	
GRIIFCIISK	1392	11	14	100	0.0003

HCY A03 Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
GGIKPAIVF	2605	11	11	78	
GGVLAALAA	1669	8	12	86	
GGVLAALAA	1669	9	12	86	
GGVLAALAA	1669	10	12	86	
GGVLPFR	32	8	13	93	0.0003
GGVLPFR	32	9	13	93	
GGVLAALAA	1618	9	12	86	
GGVLAALAA	1333	9	14	100	
GGVLPFR	3037	8	11	79	
GGVLPFR	1552	8	13	93	
GGVLPFR	1552	11	12	86	
GGVLPFR	1552	11	11	79	
GGVLPFR	1004	0	11	79	
GGVLPFR	968	0	11	79	
GGVLPFR	2921	0	11	79	0.0100
GGVLPFR	2921	10	11	79	
GGVLPFR	1782	10	14	100	
GGVLPFR	1569	8	13	93	
GGVLPFR	1569	9	13	93	
GGVLPFR	1238	10	12	86	
GGVLPFR	1238	11	12	86	
GGVLPFR	1238	10	12	86	
GGVLPFR	1131	11	12	86	
GGVLPFR	1131	11	12	86	
GGVLPFR	1131	8	11	79	
GGVLPFR	2641	0	11	79	
GGVLPFR	2063	10	14	100	
GGVLPFR	1335	10	12	86	
GGVLPFR	1863	8	12	86	
GGVLPFR	1863	9	12	86	
GGVLPFR	1863	10	12	86	0.3900
GGVLPFR	1863	8	11	79	
GGVLPFR	1193	8	11	79	
GGVLPFR	1081	10	11	79	
GGVLPFR	1081	10	11	79	0.0014
GGVLPFR	3035	10	11	79	
GGVLPFR	1670	8	12	86	
GGVLPFR	1670	9	12	86	0.0046
GGVLPFR	1670	11	11	79	
GGVLPFR	45	11	14	100	
GGVLPFR	2619	9	14	100	
GGVLPFR	2619	11	12	86	
GGVLPFR	2619	11	11	79	
GGVLPFR	154	9	11	79	
GGVLPFR	1900	10	11	79	
GGVLPFR	1900	11	11	79	
GGVLPFR	1900	11	13	93	
GGVLPFR	33	11	13	93	
GGVLPFR	33	8	11	79	
GGVLPFR	1141	8	11	79	
GGVLPFR	1141	9	11	79	

HCV A03 Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
HADVPRHRR	1141	10	11	79	
HAPTGSKG	1234	0	14	100	
HAPTGSKGSTK	1234	11	13	93	
HGI SAFSUI	2920	9	11	79	
HGI SAFSLISY	2920	11	11	79	
HGPILLY	1624	8	11	79	
HGPILLYR	1624	9	11	79	
HIDAHL SOTK	1572	11	14	100	
HLHAPTGSKG	1232	10	12	86	0.5900
HUKXNDVOY	696	11	11	79	
HLHCHSK	1395	0	14	100	
HLHCHSKK	1395	9	14	100	0.0260
HLHCHSKKK	1395	10	14	100	0.0260
HMMNHLSGIY	1769	11	13	93	
HSKKKDELA	1400	10	14	100	
HSKKKDELA	1400	11	14	100	
HSSYSGEINR	2928	10	11	79	0.0004
HTPGCVPCVR	222	10	11	79	
HMGHSEGA	1910	0	11	79	
IAFASTGNH	1925	9	14	100	0.0003
IDAHFLSOTK	1573	10	14	100	
IDHLCGFA	123	8	12	86	
IDHLCGFA	123	9	12	86	
ILGHSKKK	1397	8	14	100	
IGTVLDOA	1334	0	14	100	
IGTVLDOA	1334	11	14	100	
IGTVLDOAETA	1334	8	12	86	
HLHDECH	1317	10	11	79	
ILAGYGAGYA	1856	11	12	86	
ILGGWVAA	1816	8	12	86	
ILGGWVAAQA	1816	11	12	86	
ILGGWVAAQA	1331	11	12	86	
IMAKNEVF	2591	8	12	86	
ISGIQYLA	1774	8	14	100	
ITWESENK	2250	9	12	86	0.0150
ITSCSSNNSVA	2816	11	14	100	
ITWGADIA	989	8	12	86	
ITWGADIAA	909	9	12	86	
ITYSTYK	1296	8	12	86	
ITYSTYKGF	1296	9	12	86	
ITYSTYKELA	1296	11	11	79	
NDVOXYLY	701	0	12	86	0.0036
WFDLGVR	2613	9	11	79	
WGVVLLPH	30	10	13	93	0.0008
WGVVLLPHR	30	11	13	93	
KALGLOTA	1736	9	12	86	

ICV A03 Motif with Binding Information

Sequence	Position	No of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
KCDLAAK	1404	8	12	86	
KFGYGAQDVR	2553	10	12	86	
KGGTLLIF	1391	8	11	79	
KGGNLLFGH	1391	10	11	79	
KGGFKPAR	2604	8	11	79	
KLGPPLR	2944	8	12	86	
KSTKVPAA	1241	8	12	86	0.0009
KSTKVPAAV	1241	9	12	86	
KSTKVPAAV	1241	10	12	86	
KSTKVPAAV	1241	11	11	79	
KSTKVPAAV	1241	11	12	86	
KIKRNINR	10	8	12	86	
KIKRNINR	10	9	12	86	0.0110
KISERSQPI	51	9	13	93	0.1600
KISETSQPIGR	51	11	12	86	
KVIDILLTCGF	121	10	12	86	
KVIDILLTCGF	121	11	12	86	
KVIDILLTCGF	121	11	14	100	
KVLVLPNSVA	1255	10	14	100	
KVLVLPNSVA	1255	11	14	100	
KVPAAYAA	1244	8	11	79	
KVPAAYAA	1305	10	11	79	
LADGGCSGGA	1305	11	11	79	
LADGGCSGGA	1305	11	12	86	
LAEQIKOK	1729	8	12	86	
LAEQIKOK	1729	9	12	86	
LAEQIKOK	1057	9	11	79	
LAGYGAGVA	1057	11	11	79	
LAGYGAGVAGA	1057	10	11	79	
LCECYDAGCA	1522	10	12	86	
LDOAETAGA	1338	9	12	86	
LIXOETAGAR	1338	10	12	86	
LILLADA	727	8	14	100	
LILLADAR	727	9	14	100	
LFINLGGWA	1813	10	12	86	
LFINLGGWA	1813	11	12	86	
LFINLGGWAA	1813	11	11	79	
LFTSPRR	290	8	12	86	0.0810
LFTSPRR	1267	9	12	86	
LFGAYMSK	1267	10	12	86	
LFGAYMSKA	1267	11	12	86	
LFGAYMSKAH	1267	9	11	79	
LGAARALA	144	10	11	79	
LGAARALAI	144	10	12	86	
LGGWMAQA	1817	10	12	86	
LGGWMAQA	1332	10	13	93	
LGGWMAQA	44	8	12	86	
LGVRAIRK	2618	8	14	100	
LGVRCCEK	2618	10	14	100	
LGVVCEKMA	2618	10	14	100	
LIAFASRGNI	1924	10	14	100	
LIEANLWIR	2235	9	12	86	0.0008

ICV A03 Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
IFCISKK	1396	8	14	100	
IFCISKKK	1396	9	14	100	0.5400
IININGSM1	414	9	11	79	
IWFPOKGVH	2612	10	11	79	0.0003
ILAPITAY	1030	8	14	100	
ILFILLADA	726	9	14	100	0.0016
ILFILLADAR	726	10	14	100	
ILFNILGGWYA	1812	11	12	86	
ILPAILSPGA	1887	10	13	93	0.0003
ILPPIRGPH	36	8	13	93	
ILSPRGSR	97	8	12	86	
LMGYPLVGA	133	10	11	79	
LSAFSLHSY	2922	9	11	79	0.0002
LSAPSLKA	2211	8	11	79	
LSNSLRAH	2479	8	12	86	
LSNSLRRH1	2479	9	12	86	0.0003
ISTGLH1H	690	9	14	100	
ISTIPGNPA	1703	9	12	86	
ITCGFADLMGY	126	11	14	100	
LTDPSSHTA	2180	9	14	100	
LTIDDAHF	1570	8	13	93	
LTSMLTDPH1	2176	10	13	93	
LVAQATVCA	1591	10	12	86	
LVAQATVCAH	1591	11	11	79	
LVDLAGY	1053	8	11	79	
LVDILAGYGA	1053	10	11	79	
LWGVLA	1667	8	12	86	
LWGVLAALA	1667	10	12	86	
LWGVLAALAA	1667	11	12	86	
LVLNPSVA	1257	8	14	100	
LVLNPSVAA	1257	9	14	100	
LWGVCA	1897	8	11	79	
LWGVCAAA	1897	9	11	79	
LWGVCAAA	1897	8	11	79	
LWICESA	2773	8	11	79	
MGFSYDTR	2668	8	11	79	
MGFSYDTRCF	2668	10	11	79	
MGSSYGOY	2640	9	11	79	
MGYPPLVGA	134	9	11	79	
MILMTHFF	2876	8	12	86	
MLTDPSHITA	2179	10	14	100	
MSTINPKPOR	1	9	11	79	
MSTINPKPOR	1	10	11	79	
NOGYRQIA	2726	8	11	79	
NOGYRQIA	2726	9	11	79	
NCSTPGH	305	8	11	79	

0.0008

HCY A03 Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
PIGSGKSTK	1236	9	13	93	0.0002
PIIIVPESDA	1936	10	12	86	
PIIIVPESDMA	1936	11	12	86	
PIIHGPIPLY	1621	11	11	79	
PIILYILGA	1626	10	13	93	
PVQGNIEF	1554	9	12	86	
PVWGTTDA	516	9	13	93	0.0008
QAEIAGAR	1340	8	12	86	
QATVCARA	1595	8	13	93	
QATVCARAQA	1595	10	11	79	
QWGGVILPR	29	11	13	93	
QLEFSPRI	209	0	12	86	
QLEFSPRI	209	9	11	79	0.7500
QLEFSPRI	209	9	11	79	
QLIHPOA	336	0	12	86	
QLSAPSILK	2210	0	11	79	
QLSAPSILK	2210	9	11	79	
QIVDSIDPIF	1465	11	12	86	
RAAVCTRGVA	1186	10	11	79	
RAAVCTRGVAK	1186	11	11	79	
RAALHGVR	149	8	14	100	
RATIKTSER	47	9	11	79	
RGINNSPTI	1930	9	12	86	0.0003
RGINVSPITIIY	1930	10	12	86	0.0003
RGINLGVN	40	0	13	93	
RGPILGVNA	40	9	13	93	
RGPILGVNATR	40	11	11	79	
RGNIOPIPK	59	9	13	93	0.0120
RGSLLPR	1154	0	12	86	
HGVAKAVDF	1192	9	11	79	
RLGVNATR	43	8	11	79	
RLGVNATRK	43	8	11	79	0.9400
RLHGLSAF	2918	8	12	86	
RLHGLSAFSLI	2918	11	11	79	
RLAFASR	1923	8	14	100	
RLAFASRGINI	1923	11	14	100	
RLVFPOLGVR	2611	11	11	79	
RLAPITA	1029	0	12	86	
RLAPITAY	1029	9	12	86	2.7000
RLVVLATA	1347	0	12	86	
RLMLMTIIF	2875	8	12	86	
RLMLMTIIF	2875	9	12	86	
RLMYGVEI	635	9	14	100	
RLMYGVEIHR	635	10	14	100	0.7200
RSQTHGSR	55	8	13	93	
RVCCKMALLY	2621	9	14	100	0.1800

161

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
FMELGVNY	156	9	12	86	0.0120
FMELGVNY	156	10	12	86	
SAFSLISY	2293	8	11	79	
SASQI SAPSLK	2207	11	11	79	
SCSSNSVYA	2018	9	14	100	
SCSSNSVYAH	2818	10	12	86	
SDL YLVIRH	1133	8	12	86	
SDL YLVIRH	1133	9	12	86	
SDL YLVIRH	1133	10	12	86	
SDL YLVIRH	1133	10	14	100	
SI SIFLLA	173	8	14	86	
SGKSTKVPAA	1239	9	12	86	
SGKSTKVPAA	1239	10	12	86	
SGKSTKVPAA	1239	11	12	86	
SGKSTKVPAA	1239	11	14	100	
SMLTDP SHITA	2170	8	14	100	
SMLTDP SHITA	2170	11	14	100	
SSASQLSA	2206	8	14	100	
SSDL YLVIRH	1132	9	12	86	0.0003
SSDL YLVIRH	1132	10	12	86	0.0003
SSDL YLVIRH	1132	11	12	86	
SSDL YLVIRH	1132	11	12	86	
SSNSVYAH	2020	8	12	86	
SSNSVYAH	2205	9	14	100	
SSASQLSA	2205	8	12	86	
STGILHLH	691	8	12	86	
STKVPAA	1242	8	12	86	
STKVPAA	1242	9	12	86	
STKVPAA	1242	10	11	79	
STKVPAA	1242	10	14	100	
SILPGNPA	1704	8	14	100	
STNPKPQRI	2	8	11	79	
STNPKPQRI	2	9	11	79	
STNPKPQRI	2	11	11	79	
STNPKPQRI	2	11	11	79	
STNPKPQRI	2	11	12	86	
STNPKPQRI	1663	11	12	86	
STNPKPQRI	1299	8	12	86	
STYKFLA	1262	8	14	100	
SVAATLGF	1262	8	14	100	
SVAATLGF	1262	10	14	100	
SVAATLGF	1262	11	14	100	
SVAATLGF	1262	10	12	86	
SVAATLGF	1343	10	13	93	
TAGARLVVLA	127	10	11	79	
TCGFADMGY	1129	8	11	79	
TCGSSDL	1129	8	12	86	
TCVLTQVDF	1461	9	12	86	
TUPTKRSR	110	8	12	86	
TUPTKRSR	2101	8	14	100	
TUPTSHITA	2101	8	11	79	
TGELPYGK	1375	10	11	79	
TGELPYGK	1375	8	13	93	
TGLTHIDA	1568	9	13	93	
TGLTHIDA	1568	9	13	93	
TGLTHIDALF	1568	10	13	93	

HCV Δ93 Motif with Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*0301
TCMFGCSF	166	9	13	93	
TGSGKSTK	1237	8	13	93	
TGSGKSTKVP	1237	11	12	86	
TIKAKNEVF	2590	9	11	79	0.0810
TLGFAYMSK	1266	10	12	86	
TLGFAYMSKA	1266	11	12	86	
TLHGPIPLY	1622	10	11	79	0.0890
TLHGPIPLYR	1622	11	11	79	
TLPALSTGLH	606	11	11	79	
TLWARMILMTH	2071	11	11	79	
TSCSSNWSVA	2017	10	14	100	
TSCSSNWSVAH	2017	11	12	86	
TSEKQPH	52	8	13	93	0.0003
ISEKQPHGHI	52	10	12	86	
TSEKQPHGHI	52	11	12	86	
TSEKQPHGHI	52	8	12	86	
TSLGRDK	1050	9	13	93	0.0003
TSLGRDK	2177	9	11	79	
TSMIDPSH	2589	10	11	79	
TIKAKNEVF	1208	8	12	86	
TIKAKNEVF	1597	8	11	79	
TVCAHQAQ	1466	10	12	86	
TVDFSLDPTF	1336	9	14	100	
TVLQAETA	1336	11	12	86	
TVLQAETAGA	1263	9	14	100	
VNATLGFQV	1064	10	14	100	
VAGALVAF	1064	9	12	86	0.2400
VAGALVAFK	1064	9	12	86	
VAYOATVCA	1592	9	12	86	
VAYOATVCAH	1592	10	11	79	0.0005
VAYOATVCAH	1592	11	11	79	
VCAAILRR	1902	8	11	79	
VCAAILRRH	1902	9	11	79	
VCEKMALY	2622	8	14	100	
VCGPVYCF	505	8	13	93	
WQDLEEF	1555	8	12	86	
VCTRGVAK	1109	8	11	79	
VCTRGVAKA	1109	9	11	79	
VCMIVHGA	1082	9	11	79	
VDFSLDPTF	1467	9	14	100	
VDLAGYGA	1854	9	11	79	
VDPYRLMH	614	9	13	93	
VDPYRLMHY	614	10	13	93	
VFCVQPER	2597	8	12	86	
VFCVQPERGHI	2597	11	11	79	
VIPDLGVR	2614	8	11	79	

[illegible]

567

Table XVII

HCY All Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*1101
AACNWTGER	647	10	12	86	0.0140
AARALAHGR	147	10	11	79	
AATLGFAY	1264	9	14	100	
AAVCIHGVAK	1187	10	11	79	
ACNWTGER	648	9	12	86	
ADGGCGGAY	1306	10	11	79	
ADVLPVRR	1142	8	12	86	
ADVLPVRR	1142	9	11	79	
AFASRGNI	1926	8	14	100	
AGALVAFK	1065	0	12	86	
AGVAGALVAFK	1062	11	12	06	
AGWLSPIH	94	0	12	06	
AGWLSPIH	94	11	12	06	
AGWLSPIH	689	0	12	06	
ALSTGLIH	609	10	12	06	0.0027
ALSTGLIH	2208	10	11	79	
ASQLSAPSIK	1928	11	12	86	
ASRGNIWSPH	1265	0	14	100	
ATLGFAY	1265	0	12	86	
ATLGFAYMSK	1265	11	11	79	
ATRKTSER	48	8	11	79	0.0250
AVCTHGVAK	1100	9	11	79	
CAALLRH	1903	0	13	93	
CGFADLMGY	120	9	13	93	
CGNLTIC	2742	0	11	79	
CGSSDIYLVIR	1130	11	12	86	
CHRLGVPIIR	2941	11	11	79	
CHKSIPGHI	304	9	11	79	
CHNWTGER	049	0	12	06	
CSSNVSVAH	2819	9	12	06	
CICGSSDIY	1128	9	11	79	0.0063
CTWMNSTGFTK	555	11	11	79	0.7500
CVQTEKGR	2599	9	11	79	0.0005
CVQTEKGR	2599	10	11	79	0.0008
DAHFLSOTK	1574	9	14	100	0.0005
DEGCGSGAY	1307	9	11	79	
DMICDECH	1316	9	12	86	
DLGVRVCEK	2617	9	13	93	0.0002
DLVLTIRH	1134	8	12	86	
DVLPVIRH	1143	8	11	79	
ECYDAGCAWY	1524	10	11	79	
EGANQMANR	1915	9	14	100	0.0014
EMGNITR	2245	8	12	86	
EVFCVPRK	2596	9	12	86	0.0270
FOVPEKGR	2598	10	11	79	
FOVPEKGRK	2598	11	11	79	

HCV A11 Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*1101
FGAYMSKAI	1269	9	12	86	
FGYAKQVIR	2554	9	12	86	0.0005
FLTLADAR	728	8	14	100	
FTEAMTHY	2792	8	14	100	
FTIGLTHIDAI	1567	10	13	93	
GAARALAH	146	8	11	79	
GAARALAHGVR	146	11	11	79	
GAQMNNIR	1916	8	14	100	
GAYMSKAI	1270	8	12	86	
GFADIMGY	129	8	13	93	
GFAYMSK	1260	0	12	86	
GFAYMSKAI	1260	10	12	86	
GTQYSYQQI	2645	9	11	79	
GGAMTALAI	145	9	11	79	
GGCSGGAY	1308	8	11	79	
GGKQNGGY	26	10	14	100	
GGKNGGY	27	9	14	100	0.0001
GGRLIFQI	1392	9	14	100	
GGRLKIQISK	1392	11	14	100	
GGVLAALAAV	1669	10	12	86	
GGVLLPR	32	8	13	93	
GGVLLPRR	32	9	13	93	0.0010
GIVLLPIN	3037	8	11	79	
GLPVCOQH	1552	8	13	93	
GLPVSAIR	1004	8	11	79	
GLSAFSLH	2921	0	11	79	
GLSAFSLHSY	2021	10	11	79	0.0005
GLTHIDAH	1569	8	13	93	
GNNSPTI	1931	8	12	86	
GNIVSPITV	1931	9	12	86	
GNIRVESENK	2248	11	12	86	
GSSDLVLTIR	1131	10	12	86	
GSSDLVLTIRI	1131	11	12	86	
GSSYGFQY	2641	8	11	79	
GTFPINAY	2063	8	11	79	
GVAGALVAFK	1863	10	12	86	1.4000
GVCAIVVI	1081	8	11	79	
GVGVILLPIN	3035	10	11	79	0.0140
GVLAALAAV	1670	9	12	86	0.0110
GVRAIRKTSER	45	11	11	79	
GVRCCKMALY	2619	11	14	100	
GVRLDQVNY	154	11	12	86	
GVVCAAILR	1900	9	11	79	
GVVCAILLIR	1900	10	11	79	
GVVCAILLIRI	1900	11	11	79	

HCV All Motif With Binding Information

Sequence	Position	No of Amino Acids	Sequence Frequency	Conservancy (%)	A*1101
GVYLLPHH	33	8	13	93	
GVYLLPHRGR	33	11	13	93	
HADVPR	1141	8	11	79	
HADVPRH	1141	9	11	79	
HADVPRHR	1141	10	11	79	
HAPTGSGK	1234	8	14	100	
HAPTGSGKSTK	1234	11	13	93	
HGLSAFSLIH	2920	9	11	79	
HGLSAFSLSHY	2920	11	11	79	
HGPILLY	1624	8	11	79	
HGPILLYR	1624	9	11	79	
HIDAFLSQIK	1572	11	14	100	
HIAPIGSGK	1232	10	12	86	0.0024
HUKKNDVOY	696	11	11	79	
HLFCISK	1395	8	14	100	
HLFCISKK	1395	9	14	100	0.0006
HLFCISKKK	1395	10	14	100	0.0002
HMMNHSGLQY	1769	11	13	93	
HSYSPGEINR	2920	10	11	79	
HTPGCVPCVR	222	10	11	79	0.0012
IATASRGNH	1925	9	14	100	0.0003
IDAFLSQIK	1573	10	14	100	
IFCHSKKK	1397	8	14	100	
MCDECH	1317	8	12	86	
INTNGSMH	415	8	11	79	0.0079
ITTVSEENK	2250	9	12	86	
ITVSTYGG	1206	8	12	86	
INDVOYLY	701	8	12	86	
IVFDIGVR	2613	9	11	79	0.0044
IVGVYLLPR	30	10	13	93	0.0056
IVGVYLLPRR	30	11	13	93	
KCDEIAK	1404	8	12	86	
KFGYGAQVHR	2553	10	12	86	
KGGHLEFCH	1391	10	11	79	
KGGKPAR	2604	8	11	79	
KLGVPILR	2944	8	12	86	
KNEYFCVQTEK	2594	11	11	79	0.0001
KSTKVPAAV	1241	9	12	86	
KIKNTNRR	10	8	12	86	
KIKNTNRR	10	9	12	86	0.0100
KISERSQPR	51	9	13	93	0.0640
KTSERSQPRGR	51	11	12	86	
LAQGGCSGAY	1305	11	11	79	
LAQRKK	1729	8	12	86	
LDQAEIAGAR	1338	10	12	86	

HCY All Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*1101
LFLLADAH	727	9	14	100	
LFESPRH	290	8	11	79	
LGFGAYMSK	1267	9	12	86	0.2900
LGFGAYMSKAH	1267	11	12	86	
LGGAARALAH	144	10	11	79	
LGVRATPK	44	8	12	86	
LGVRCEK	2618	8	14	100	
LIASFAGNH	1924	10	14	100	
LIENILMR	2235	9	12	86	0.0005
LIFCHSKK	1396	8	14	100	
LIFCHSKKK	1396	9	14	100	0.1900
LINTGSMH	414	9	11	79	
LIVEPOLGVR	2612	10	11	79	0.0001
LIAPITAY	1030	8	14	100	
LIPELLADAH	726	10	14	100	
LIPIHGRH	36	8	13	93	
LIPIHGRH	97	8	12	86	
LSAFSLHSY	2922	9	11	79	0.0002
LSNQLLNI	2479	8	12	86	
LSNQLLNIH	2479	9	12	86	0.0001
LSNQLLNIH	690	11	12	86	
LSNQLLNIH	126	11	12	86	
LTGCFADLMGY	2176	10	13	93	
LTSMALDPSH	1591	11	11	79	
LVAYQATVCAR	1853	8	11	79	
LVQILAGY	2660	8	11	79	
MGFSYDTH	2640	9	11	79	
MGSSYGQY	1921	10	14	100	
MNRLIAFASR	558	8	11	79	
MNSTGFTK	1	9	11	79	
MSTNPKPQR	1	10	11	79	
MSTNPKPQR	2726	8	11	79	
NCGYRQR	305	8	11	79	
NCSYRQH	1772	8	14	100	
NFSGQY	1080	8	11	79	
NGVCMIVY	1080	9	11	79	
NGVCMIVH	2249	10	12	86	0.0062
NITRVESENK	700	9	12	86	0.0140
NIVDOYLY	14	10	11	79	0.0007
NINRIRPOLVK	1549	11	13	93	
NIPQLPVCOOH	608	9	12	86	
PALSTGLIH	688	11	12	86	
PALSTGLIHH	1976	8	11	79	
PCSGSMR	1127	10	11	79	
PCGSSQY	2616	10	13	93	
PDGVRCEK		10			

ICV A11 Motif With Binding Information

Sequence	Position	No of Amino Acids	Sequence Frequency	Conservancy (%)	A*1101
PGQPCVH	224	8	12	86	
PGEGAVQMAN	1913	11	13	93	
PGQGVGVY	25	11	14	100	
PGAPVQDHI	1551	9	13	93	
PGYPMPLY	79	8	14	100	
PIYSTYGK	1295	9	11	79	
PIGGAARALAH	143	11	11	79	
PMGFSYDTR	2667	9	11	79	
PMRTGVR	1281	8	13	93	
PSPVVGITDR	514	11	13	93	
PSWQAMK	1607	0	11	79	
PTDCTIKI	587	0	13	93	
PIDPPIHNSR	109	9	12	86	0.0005
PTGSGSKIK	1236	9	13	93	0.0001
PIIHGPIPLY	1621	11	11	79	
PVVGITDR	516	9	13	93	0.0005
QAEIAGAR	1340	8	12	86	
QWGVYII PR	29	11	13	93	
QLFISPR	289	8	12	86	
QLFISPRH	289	9	11	79	0.0330
QLSAPSLK	2210	0	11	79	
QNVDMOY	699	0	11	79	
QNVDMOYL	699	10	11	79	
QNAVCTHGVAK	1106	11	11	79	
PAALIGVR	149	0	14	100	
PAIRKTSER	47	9	11	79	
RGNMNSPTI	1930	9	12	86	0.0001
RGNMNSPTIY	1930	10	12	86	0.0001
RCPRLGVR	40	8	13	93	
RCPRLGVIATR	40	11	11	79	
RCPRRCPRK	59	9	13	93	0.0017
RGSLSPR	1154	8	12	86	
RLGVRAIR	43	8	11	79	
RLGVRAIRK	43	9	11	79	0.0290
RLHGLSAFSLH	2918	11	11	79	
RLIAFASR	1923	8	14	100	
RLIAFASRGNH	1923	11	14	100	
RLMFPDLGVR	2611	11	11	79	
RLALPITAY	1029	9	12	86	0.0270
RLMVGVEII	635	9	14	100	
RLMVGVEIIR	635	10	14	100	0.0200
RNINIRPODMK	13	11	11	79	
RSORRRR	55	0	13	93	
RVCCKMAY	2621	9	14	100	
RMLEGVNY	156	9	12	86	0.0068

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*1101
SASFILISY	2923	8	11	79	
SASQI SAPSLK	2207	11	11	79	
SCSSNVSAH	2818	10	12	86	
SDL YLVIRH	1133	8	12	86	
SDL YLVIRH	1133	9	12	86	
SGSKTKVPAAY	1239	11	12	86	
SMLTDP SH	2178	8	14	100	
SNKLRRHH	2480	8	12	86	0.0044
SSDL YLVIRH	1132	9	12	86	
SSDL YLVIRH	1132	10	12	86	0.0013
SSNVSVAH	2020	8	12	86	
STGLHLH	691	8	12	86	
STKVPAAY	1242	8	12	86	
STKPQR	2	8	11	79	
STNPQR	2	8	11	79	
STNPQR	2	9	11	79	
STNPQR	2	11	11	79	
SVATLGF GAY	1262	11	14	100	
TGCFALMGY	127	10	13	93	
TGGSSDLY	1129	8	11	79	
TDPRRSR	110	8	12	86	
TGEIPFYGK	1375	9	11	79	
TGLTHIDAH	1568	9	13	93	0.0001
TGSGKSITK	1237	8	13	93	
TLGFAYMSK	1266	10	12	86	0.0610
TLHGPIPLY	1622	10	11	79	0.0007
TLHGPIPLY	1622	11	11	79	
TLPALSTGLH	600	11	11	79	
TLWAMLMITH	2871	11	11	70	
TNPKQR	3	8	11	79	
TNPKPQRITK	3	10	11	79	
TNPKPQRITKR	3	11	11	79	
TNRPQDYK	15	9	11	79	
TSCSSNVSAH	2817	11	12	86	
TSESRQR	52	8	13	93	
TSENSORQR	52	10	12	86	0.0001
TSETSORNGRH	52	11	12	86	
TSLTGDK	1050	8	12	86	
TSMLTDP SH	2177	9	13	93	0.0001
VAAITLGF GAY	1263	10	14	100	
VAGALVAFK	1864	9	12	86	0.8900
VAYQATVCAR	1592	10	11	79	0.0038
VCAAILRH	1802	8	11	79	
VCAAIL RH	1902	9	11	79	
VCEKMAL Y	2622	8	14	100	
VCTINGVAK	1189	8	11	79	

HCV All Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*1101
VDYPTILWHL	614	9	13	93	
VDYPTILWHL	614	10	13	93	
VCQVQNEK	2597	8	12	86	
VFQVQNEKCH	2597	11	11	79	
VFPQGVH	2614	8	11	79	
VFTGLIHDAL	1566	11	13	93	
VGGVLAALAY	1668	11	12	86	
VGGVLLPR	31	9	13	93	0.0019
VGGVLLPRH	31	10	13	93	
VGNLLPRH	3036	9	11	79	0.0100
VGVVCAILLH	1099	10	11	79	
VGVVCAILLH	1099	11	11	79	
VLAALAY	1671	8	12	86	
VLDIAETAGAH	1337	11	12	86	
VEDGVNY	157	8	12	86	
VLTSMILDPH	2175	11	13	93	
VLDVLAGY	1852	9	11	79	
VMSSTGYOY	2639	10	11	79	
VTHIDVLPVR	1138	11	11	79	
VVCALLH	1901	8	11	79	
VVCALLH	1901	9	11	79	
VVCALLRH	1901	10	11	79	
VGVVCAILLH	1098	11	11	79	
VVGITDIL	517	8	13	93	
VAGWLSPIH	93	9	12	86	
WAPGYPWPL	76	11	12	86	
WARMILMIL	2073	9	12	86	
WGPTDPRH	107	8	12	86	
WGPTDPRH	107	9	12	86	
WGPTDPRH	107	11	12	86	
WGPTDPRH	107	11	12	86	
WLSPIKSR	96	9	12	86	0.0005
WMNLRLAFASR	1920	11	14	100	
WMNSTIGETK	557	9	11	79	0.0810
WNFSGIQY	1771	9	14	100	
YDAGCAWY	1526	8	11	79	
YDILCDECH	1315	10	12	86	
YGFQYSPQCH	2644	10	11	79	
YLPKQFTH	35	9	13	93	0.0005
YSPGEINL	2930	8	11	79	
YVGVEIRH	637	8	14	100	
YVPESDAAAH	1939	10	12	86	0.0001
311		3			

Table XVIII

HCY A24 Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*2401
AWDMAMNW	319	8	12	86	
AYAGGYKVL	1248	10	11	79	0.0009
AYYHGLDVSIL	1421	11	14	100	
CYDAGCAW	1525	8	11	79	
CYDAGCAWYEL	1525	11	11	79	
DFSLDPITF	1468	8	14	100	
DFSLDPITFI	1468	10	14	100	6.9000
FWAGIMMWF	1765	9	12	86	
FWAGIMMWF	1765	10	12	86	
GFADLMGYI	129	9	13	93	
GFADLMGYPL	129	11	11	79	
GFADLMGYPL	129	9	11	79	
GFSDTRCF	2669	8	11	79	
GWILLAP	1027	8	11	79	
GYGAGVAGAL	1859	10	12	86	0.0003
GYPLVGAPL	135	10	11	79	0.0057
GYHRCASGM	2728	11	12	86	
IMMNFISGL	1769	9	13	93	
IELLALLSCL	176	10	12	86	
IMAKNEVF	2591	8	12	86	
KFGGQGL	23	8	13	93	
LINILGW	1813	8	12	86	
LWATMLMTHF	2872	11	12	86	
LWROEGSN	2241	10	12	86	
LYLVTRHADV	1135	11	11	79	
MMNFISGL	1770	8	14	100	
MMNFISGL	1770	11	14	100	
MMNFISGL	1770	10	13	93	0.0270
MMNGVEHFL	636	9	14	100	0.0170
NFISGL	1772	9	14	100	
PMGFSDTRCF	2667	11	11	79	
QFKOKALGL	1732	9	12	86	
QFKOKALGL	1732	10	12	86	
OMMNHUAF	1919	9	14	100	
OYLALSTL	1778	9	14	100	0.0480
OYSGQVEF	2647	10	11	79	0.0180
OYSGQVEFL	2647	11	11	79	
FMAMDMAMNW	317	10	12	86	
FMILMTHF	2875	8	12	86	
FMILMTHF	2875	9	12	86	
FMVGGVEHFL	635	11	13	93	
SFSLIAL	173	9	14	100	0.0041
SFSLIAL	173	10	14	100	
SMLTDPISH	2178	9	14	100	
SWDOMAKOL	1608	11	11	79	
SYKSSSGRL	1164	11	12	86	
TVANSTGF	556	8	11	79	

ICV A24 Motif With Binding Information

Sequence	Position	No. of Amino Acids	Sequence Frequency	Conservancy (%)	A*2401
TMELVGVL	1664	9	12	0.6	
TYSTYGKF	1297	8	13	93	
TYSTYGKFL	1297	9	12	86	0.0230
VFGLTH	1566	8	13	93	
VMSSSYGF	2639	8	11	79	
VYLLPRKPRLL	34	11	13	93	0.0016
WMNHILAF	1920	8	14	100	
YYRGLDYSVI	1422	10	14	100	
53		2			

Table XIX a

ICV DR-Super Motif

Core Sequence	Core Frag	Core Consistency (%)	Exemplary Sequence	Position In ICV Poly protein	Exemplary Sequence Frequency	Exemplary Sequence Consistency (%)
IGVWKSNI	12	86	TIIGVWKSNIYVD	1266	5	36
IGVWKSNI	12	86	GVWIGVWKSNI	550	11	79
IGVWKSNI	12	86	ACVIGVWKSNI	1730	12	86
IGVWKSNI	12	86	FSIIGVWKSNI	174	6	43
IGVWKSNI	12	79	LVWIGVWKSNI	2612	11	79
IGVWKSNI	12	86	POVIGVWKSNI	1225	6	43
IGVWKSNI	12	86	VOVIGVWKSNI	1182	7	50
IGVWKSNI	12	100	GVWIGVWKSNI	171	12	86
IGVWKSNI	14	100	GVWIGVWKSNI	1466	11	79
IGVWKSNI	14	100	GVWIGVWKSNI	2789	7	50
IGVWKSNI	13	93	GVWIGVWKSNI	509	13	93
IGVWKSNI	11	79	GVWIGVWKSNI	681	9	64
IGVWKSNI	12	86	GVWIGVWKSNI	1762	3	21
IGVWKSNI	14	100	GVWIGVWKSNI	1570	7	50
IGVWKSNI	12	86	GVWIGVWKSNI	1454	12	86
IGVWKSNI	12	83	GVWIGVWKSNI	120	7	50
IGVWKSNI	12	85	GVWIGVWKSNI	2233	12	86
IGVWKSNI	14	100	GVWIGVWKSNI	173	6	43
IGVWKSNI	12	86	GVWIGVWKSNI	1813	8	57
IGVWKSNI	12	79	GVWIGVWKSNI	1328	8	57
IGVWKSNI	13	93	GVWIGVWKSNI	1903	11	79
IGVWKSNI	12	86	GVWIGVWKSNI	1888	11	79
IGVWKSNI	13	93	GVWIGVWKSNI	2064	8	57
IGVWKSNI	12	86	GVWIGVWKSNI	134	10	71
IGVWKSNI	11	79	GVWIGVWKSNI	2247	10	71
IGVWKSNI	12	86	GVWIGVWKSNI	2813	11	79
IGVWKSNI	14	100	GVWIGVWKSNI	2610	11	79
IGVWKSNI	12	86	GVWIGVWKSNI	1669	8	57
IGVWKSNI	12	79	GVWIGVWKSNI	1302	10	71
IGVWKSNI	14	100	GVWIGVWKSNI	1777	14	100
IGVWKSNI	11	79	GVWIGVWKSNI	1854	10	71
IGVWKSNI	12	86	GVWIGVWKSNI	1346	8	57
IGVWKSNI	12	86	GVWIGVWKSNI	1468	5	36
IGVWKSNI	12	86	GVWIGVWKSNI	1305	12	86
IGVWKSNI	13	93	GVWIGVWKSNI	2810	13	93
IGVWKSNI	12	86	GVWIGVWKSNI	1655	4	29
IGVWKSNI	14	100	GVWIGVWKSNI	724	8	57
IGVWKSNI	12	86	GVWIGVWKSNI	1814	9	64
IGVWKSNI	13	93	GVWIGVWKSNI	1329	10	71
IGVWKSNI	12	86	GVWIGVWKSNI	41	10	71
IGVWKSNI	14	100	GVWIGVWKSNI	2615	11	79
IGVWKSNI	11	79	GVWIGVWKSNI	2916	6	43
IGVWKSNI	11	79	GVWIGVWKSNI	1620	11	79
IGVWKSNI	12	86	GVWIGVWKSNI	694	10	71
IGVWKSNI	11	79	GVWIGVWKSNI	2924	11	79
IGVWKSNI	14	100	GVWIGVWKSNI	1921	12	86
IGVWKSNI	12	86	GVWIGVWKSNI	2232	7	50
IGVWKSNI	14	100	GVWIGVWKSNI	1393	14	100
IGVWKSNI	14	100	GVWIGVWKSNI	2812	13	93
IGVWKSNI	12	86	GVWIGVWKSNI	175	5	36
IGVWKSNI	14	100	GVWIGVWKSNI	723	5	36
IGVWKSNI	12	86	GVWIGVWKSNI	1809	4	29
IGVWKSNI	13	93	GVWIGVWKSNI	726	9	64
IGVWKSNI	13	93	GVWIGVWKSNI	1884	10	71

ICV DR-Super Model

Cure Sequence	Cure Freq	Cure Consistency (%)	Exampley Sequence	Position In ICV Poly protein	Exampley Sequence Frequency	Exampley Sequence Consistency (%)
LMKRYVQ	11	79	FAQKADYVWGNL	130	11	79
LMPSVAIL	14	100	WLVNPSVAILQFG	1256	14	100
LPALSPGA	13	53	WHLPLSPGALV	1885	11	79
LPALSTGI	12	66	FTTLPALSTGIH	684	11	79
LPSTETIG	13	93	VLLTTPKSPRIGVIA	341	13	93
LPOLAVVE	11	79	HKKLTOLAVVEPV	966	4	29
LPRLGVPL	12	66	ASCLNLOVPLTW	2939	7	50
LSVSLHSY	11	79	ILDLVSLHSYSG	2919	11	50
LSAPSKAT	11	79	ASQKSNPSKATCT	2208	7	29
LSNSLNRH	12	86	HNLSNSLNRHRAV	2476	4	29
LSFGALVQ	13	93	PAISFGALVGVWC	1889	11	79
LSPLLSIT	11	79	NSELPLLSITTEWQ	664	7	50
LSREKTS	11	79	GMALSRTEKTSMA	95	11	79
LSGLHRL	12	66	LPALSGLHRLDIA	607	10	71
LIQCFALM	12	66	IKDLIQCFALMST	123	12	86
LIHIDNHL	13	93	FTDLIHIDNHLSDI	1567	13	93
LIEMKTPS	13	93	VAVLISMALDPSH	2173	9	64
LIAYOATVC	12	86	FPYLVAYOATVCNIA	1506	9	64
LIOLAGYQ	11	79	GYLVOLAGYQAGV	1850	9	64
LIQGYLAL	12	86	TWLVQGYLALNAV	1664	12	88
LIHPSVAA	14	100	YKVLHPSVAAITG	1254	14	100
LIHILPAL	11	79	TEHVLHILPALSHQ	1881	10	71
LIYTRINDVI	11	79	DLVLYTRINDVPIV	1134	11	79
LIQGVVCA	11	79	PIVALQGVVCAHII	1094	11	79
LIYLATIP	12	86	QMLIYLATIPFOS	1345	11	79
LIWIRHML	12	86	APYLIWIRHMLHFE	2069	11	79
LIWKEKXZH	11	79	AKLLWKEKXZHII	2236	12	86
LIYHIGAVCH	12	79	HTLYHIGAVCHV	1627	9	64
LMKREVCV	12	86	THMKREVCVQRE	2509	9	64
LMVNMAMAM	12	86	CTLMVNMAMAMNPT	315	12	88
LMKALINE	11	79	PKMKALINEEHI	2243	11	79
LMYHLYOA	14	100	AKLMYHLYOANPQ	131	8	57
MLTDSHIT	14	100	UTSMLTDSHITAL	2176	14	100
MLHILNAS	14	100	VDMMHILNASQKSI	1918	14	71
MLTTSVTO	14	100	TEMLTTSVTPKMP	2793	12	86
MMNISJQ	14	100	AKHMMNISJQDIA	1767	12	36
MMKQVEH	14	100	KHMMKQVEHRIIA	633	5	50
VAGALVAFK	12	86	GAGVAGALVAFKMS	1861	7	43
VAILHAPTG	12	86	TIOVALHAPTGSK	1227	6	43
VAILDALMG	12	86	VWVALDALMGITG	1437	6	43
VAYOATVCA	12	86	PYLVAYOATVCANQ	1589	6	79
VCMALDRIH	14	79	GVVCMALDRIHGP	1899	11	71
VCEHMAFY	12	86	GVINCEHMAFYVWS	2619	10	79
VCOKEEFW	12	86	GLPVCOKEEFWESV	1552	6	43
VCHTGVNA	11	79	HNACHTGVNAKVP	1186	11	79
VFCVDEKQ	12	86	HEMFCVDEKQKIK	2594	10	71
VFTKSSPP	11	79	NSPVFTKSSPPVP	1211	10	71
VFTGLTHD	13	93	WESVFTGLTHDAFE	1563	6	43
VGVVIALA	12	86	WLVGVVIALANAVC	1665	12	86
VGVVILPPI	13	93	GOVGVVILPPIKXP	28	13	93
VGSATCEP	12	86	OTLVGSATCEPERD	2158	6	43
VGVVCAIL	11	79	ALVGVVCAILRIH	1896	11	79
WIKCHICVT	12	86	FDSWIKCHICVOTIV	1453	12	86
WDLTICGF	12	86	LQKWDLTICGFAL	119	11	79

HCV DR-Super Motif Binding Data Not Included

Exemplary Sequence	Position in HCV Poly protein	Exemplary Sequence	Conservancy (%)	Conservancy (%)
VGWVLAIAAYCLIT	1668	VGWVLAIAAYCLIT	86	12
RLVLAIAIPRSSVI	1347	RLVLAIAIPRSSVI	93	13
GVRLIEKVAATGN	154	GVRLIEKVAATGN	86	12
KVLVNPVSAATLGF	1255	KVLVNPVSAATLGF	100	14
DVAVLTSALIDPSIH	2172	DVAVLTSALIDPSIH	93	13
ASGVLTSTGNITC	2734	ASGVLTSTGNITC	79	13
LGKLVIMAGYAG	1849	LGKLVIMAGYAG	79	13
STWLVGVVLAIAA	1663	STWLVGVVLAIAA	86	12
GKVLVNPVSAATL	1253	GKVLVNPVSAATL	100	14
EDVNLIPALSPGA	1882	EDVNLIPALSPGA	86	12
THVPESDAARVTO	1937	THVPESDAARVTO	86	12
LEVISTWLVGVVL	1658	LEVISTWLVGVVL	86	12
DVWVAIDALMGYT	1436	DVWVAIDALMGYT	79	13
VGVVCAILRRNG	1898	VGVVCAILRRNG	79	13
GALVGVVCAILRRN	1895	GALVGVVCAILRRN	79	13
AILVLAIAIPSSV	1346	AILVLAIAIPSSV	93	13
CGVVCFTSPVWG	506	CGVVCFTSPVWG	86	12
GVGVVCAILSTRGN	90	GVGVVCAILSTRGN	86	12
PTLVARNIMHFTS	2870	PTLVARNIMHFTS	86	12
ILVGNADTACSDIL	988	ILVGNADTACSDIL	86	12
TPSKATLQNTGSGN	104	TPSKATLQNTGSGN	86	12
AVDMMNLIAFASNG	1917	AVDMMNLIAFASNG	100	14
SKGWRLAPLVAAG	1025	SKGWRLAPLVAAG	79	13
SYTNGALITPCAE	2456	SYTNGALITPCAE	86	12
GCAMVELPAETVR	1529	GCAMVELPAETVR	86	12
GVNVAIGALPGCS	161	GVNVAIGALPGCS	93	13
GVNVAIGALPGCS	507	GVNVAIGALPGCS	79	13
CEVDAGAWYELIP	1523	CEVDAGAWYELIP	86	12
GGAVDMLKECHST	1312	GGAVDMLKECHST	93	13
QREYVLEITSGSN	2808	QREYVLEITSGSN	86	12
LAGVGVVCAILVAF	1857	LAGVGVVCAILVAF	79	13
GSSVCGSTGVONE	2641	GSSVCGSTGVONE	86	12
YSTVGNALKLSLS	1298	YSTVGNALKLSLS	79	13
AGGKRVVNPVSA	1251	AGGKRVVNPVSA	100	14
GIQYLAALSNPGR	1776	GIQYLAALSNPGR	100	14
PSYTIKSSSGRLC	1162	PSYTIKSSSGRLC	86	12
IRVYVIRIPIETLRI	2833	IRVYVIRIPIETLRI	79	13
LVAYDAIVCARAQA	1591	LVAYDAIVCARAQA	93	13
VAVYRGLDVSVIPS	1420	VAVYRGLDVSVIPS	100	14
FLVYRI GAVONEVLI	1628	FLVYRI GAVONEVLI	79	13
NDGVNRCTMSGVIT	2726	NDGVNRCTMSGVIT	93	13
GACYSIGTIDPOL	2902	GACYSIGTIDPOL	79	13
LIHSYSPGKGNVASC	2927	LIHSYSPGKGNVASC	79	13
SAWVVRIRCGSVIV	273	SAWVVRIRCGSVIV	86	12
	3036		79	13

174

HCY DR Super Motif With Binding Data

[illegible]

[illegible]

HCV DR Super Motif With Binding Data

Cone Sequence	Exemplary Sequence	UK1	UK2w2 1	UK2w2 2	UK3	UK4w4	UK4w5	UK5w11	UK5w12	UK6w19	UK8w2	UK7	UK9	UK1w1
VI EGVVRA	QVINEGVVRA	0.0007				0.0006						-0.0002		
VI HPSSVAT	KVLVIHPSSVATLOF													
VI TSMILTP	UVAVITSMILTPSH													
VI TISQHT	ASQVITISQHTIC													
VI VDIAGY	LGKLVMDIAGYAG													
VI VGGVAA	STWLVGGVAAALAA													
VI VLPVYA	GYKVLVLPVYAAL	1.0000	0.0260	0.0004	0.0980	0.6000	0.0670	0.1400	0.0520	0.6900	0.1700	0.2800	0.2800	1.4000
VI LIPVLS	EELVILIPVLS	0.3700				0.0110						0.0015		
VPESDAAN	HHVPESDAANVTO													
VISTWLVG	LEVISTWLVGVVL	0.0120	0.0078	0.0003	0.0003	0.0280		0.0008		0.0046		0.1600	0.1600	0.0120
VVAIDVMT	DVVVAIDVMTGYT	0.0110	0.0110	0.0003		0.0180	0.0072	-0.0004		0.0140	-0.0003	0.0910	-0.0025	
VVCALIR	VGVVCALIRBING					0.0067						0.0043		
VGVVCAAI	GALVGVVCAAILN	0.0170												
VVIATIRP	ANVVIATIRPQSV					0.2600	0.4000	0.0005		-0.0001	0.0011	0.2700	0.4300	
VVCFTPSV	GCVCFTPSVWVG	0.2700	0.0025	-0.0003										
WAGWLSPI	GCCWAGWLSPIGSH					0.0700						0.0190		
WABWLMII	PLWABWLMIIIFS	0.0064												
WQADIMAG	HWQADIMAGSH													
WAPLUTRN	HWAPLUTRNSTN													
VANRIMFA	AVONRIMFASQ	2.2000				0.0015						0.0205		
WILLAPIA	SKGMWILLAPIAVAO	14.0000	0.0130	0.0800	-0.0006	2.1000	0.2500	4.2090	0.0290	-0.0001	0.9000	0.0280	0.0610	
WIGALITC	SYWIGALITPCAE	0.0260	0.0007	0.0015		0.0680	0.0220	0.0011		-0.0001	0.0130	0.4900	0.0750	
WHELTPAET	GCWHELTPAETYN					0.0110						-0.0003		
YAIQALFC	GMVYAIQALFCS	0.0011												
YCFITSPV	GNVCFITSPVWGT													
YDGACWYE	CECYDGACWYELP													
YDKLJEC	QDAYDKLJECIST					0.0004						-0.0002		
YDKELTSC	QFEYDKELTSCSN	0.0003				-0.0003						0.0008		
YGAQVAGAL	LAQYGAQVAGALVAF	0.0410				0.1200	0.0510	0.0010	0.0003	0.1800	0.0007	0.1600	1.1000	
YGGVGTQ3	OSSVYGGVGTQ3VE	0.4600	0.0001	0.0100	0.0007									
YGGVLTQ3	YSTVYGGVLTQ3SD													
YKVLVLPVS	AGQYKVLVLPVSAA	0.6100	0.0140	0.0004	0.0045	6.3000	0.1700	0.2700	0.0370	0.5900	0.2800	0.0300	0.2000	
YLAGI STIP	GQVYLAGI STIPAP													
YKSSXEP	PVSXKSSXEPNLC													
YLTIRPTIP	RNYLTIRPTIPTHAN													
YQATVCAIA	LVAQATVCAIACAP													
YRKLIVSMA	VAYRKLIVSMAPIIS													
YRIGAVME	PILYRIGAVMEVIL													
YRIGVGVN	HCYRIGVGVNMTT													
YSEHCLP	GACYSSEHCLPDI													
YSGENIV	UESYSGENIVASC													
YVGLX COSV	SMATYVGLX COSVTV				0.0017									
YGRILTRFI														

Table XXa

HCV DR 3A Motif Binding Data Not Included

Core Sequence	Core Freq	Core Conservancy (%)	Exemplary Sequence	Position in HCV Poly protein	Exemplary Sequence Frequency	Exemplary Sequence Conservancy (%)
ILNKKXCSA	11	79	YGRILNKKXCSGAY	1301	10	71
FSIDPFTI	14	100	INDFSIDPFTIETT	1406	11	79
ILEGEKSD	14	100	MPYLEGEKSDKSD	2401	11	79
IPCPEPDV	12	86	GSQIPCPEPDVAVL	2162	9	64
MMWDMKMMW	12	86	GIIDMMWDMKMMASPT	315	12	86
MLTDPSTLT	14	100	LTSMALTDPSTLTET	2176	8	57
MSADLEVT	11	79	MACMSADLEVTSTW	1651	6	43
VAIDALMIG	12	86	VVVVAIDALMIGYTG	1437	6	43
VCCDRLETW	12	86	GLPVCCDRLERWESV	1552	8	43
VFPDLGVIV	11	79	FLNFTDLGVIVCEK	2611	11	79
VITDKSSRP	11	79	NSPVITDKSSRPVNP	1211	10	71
VIECYDAG	13	93	DSSVLCIECYDAGCAW	1510	12	71
VLEDGVNYA	12	86	GVNMLEDGVNYAIGH	154	10	80
VLVDLAGY	11	79	LKMLVDLAGYGAG	1049	10	71
WQDQKXIK	11	79	WICQDQKXIKAPNI	2597	11	79
YIMELIISC	10	93	QPEYIMELIISCSN	2808	11	79
YSIEPLIIP	11	79	GACYSIEPLIIPDIL	2902	8	43
YVGDKCSV	12	80	SNMYVGDKCSVITV	273	8	57
YVPESDMAA	12	80	PTIIVVPESDMAAIVT	1938	12	86

19

Table XXb

HCY DR 3A Motif With Binding Information

Core Sequence	Exemplary Sequence	DR0	DR1	DR2w201	DR2w202	DR4w4	DR4w15	DR5w11	DR5w12	DR6w19	DR7	DR8w2	DR9	DRw33
FLAXXAXA	YQATLAKKXCSGAY		0.0001			0.1600					0.0005			
FSIDPTII	TVDFSLWTFIET	-0.0017												
IEGERKJND	MPTEGETGIDTK SD													
LIICEPEIIV	GSQIICEPEIDVAVL	-0.0017	0.0203	0.0015	0.0044	0.1600		0.0079		0.0080	0.0017		0.0230	
MAWMDMAWAW	GIHMAWMDMAWAWSPIT		0.0004			0.0740					-0.0003			
MA TOPSHIT	ITSMI TOPSHITAEI													
MSADLEVI	MACMSAKI EVVTSIW							0.0006		0.0029	0.0400	0.0029		
VATDAIMIG	VVVVATDAIMIGYIG	1.1000	0.0043	0.0047	0.0014									
VCCQRIELW	GI PVCCRIELFWESV	0.0063												
VFFIA GVIV	FMNFFIA GVIVNCEK													
VFTDKSSPP	RSPVFTDKSSPPAVP													
VCECYVDAG	DSSALCECYVDAGAV	-0.0017									-0.0002			
VLEDGVVYA	GVIMLEDGVVYVIGH		0.0007			0.0006								
VLVDIL AGY	LGKVLVDILAGYAG													
WCTKGAIR	WCVCTKGAIRVYNI													
YAEIITSC	QREYDAEIIISCSN		0.0003			0.0004					-0.0002			
YSIEPIIAP	GACYSIEPIIAPQII	-0.0017												
YWGICQISV	SMAYWGICQISVTLV	0.0220												
YVPESDAAA	PIIIVVPESDAAAIIVT													
19														

Table XXC HCV JB Motif

Core Sequence	Core Freq	Core Consistency (%)	Exemplary Sequence	Position in HCV Poly-protein	Exemplary Sequence Frequency	Exemplary Sequence Consistency (%)
IQEKKKK	14	100	IAIFGISEKKKKDELA	1395	14	100
FSIDHCTD	11	79	PKAUSTDHFQDSTV	2667	11	79
LAEDFKKKA	12	86	GMGLAEDFKKKAQL	1726	8	57
LVPTUHKPT	11	79	LVRLPTUHKPTPL	1616	10	71
VAATIKISE	11	79	ALGVNATIKISEED	43	10	71
YLVTRHADV	12	86	SDI YLVTRHADVIV	1133	11	79
MSLTPKTKR	11	79		1		

Table XXXd

HCY 3B Motif Binding Data

[illegible]

TABLE XXI. Population coverage with combined HLA Supertypes

<u>HLA-SUPERTYPES</u>	<u>PHENOTYPIC FREQUENCY</u>					
	Caucasian	North American Black	Japanese	Chinese	Hispanic	Average
<u>a. Individual Supertypes</u>						
A2	45.8	39.0	42.4	45.9	43.0	43.2
A3	37.5	42.1	45.8	52.7	43.1	44.2
B7	38.6	52.7	48.8	35.5	47.1	44.7
A1	47.1	16.1	21.8	14.7	26.3	25.2
A24	23.9	38.9	58.6	40.1	38.3	40.0
B44	43.0	21.2	42.9	39.1	39.0	37.0
B27	28.4	26.1	13.3	13.9	35.3	23.4
B62	12.6	4.8	36.5	25.4	11.1	18.1
B58	10.0	25.1	1.6	9.0	5.9	10.3
<u>b. Combined Supertypes</u>						
A2, A3, B7	83.0	86.1	87.5	88.4	86.3	86.2
A2, A3, B7, A24, B44, A1	99.5	98.1	100.0	99.5	99.4	99.3
A2, A3, B7, A24, B44, A1, B27, B62, B58	99.9	99.6	100.0	99.8	99.9	99.8

Table XXII

ICV ANALOGS

AA	Sequence	Fixed Normen.	A1 Moult	A2 Super Moult	A3 Super Moult	A24 Moult	B7 Super Moult	1° Anchor Fluxor
9	HVXEKMALY		N	N	Y	N	N	Nb
9	AVXTRGVAK		N	N	Y	N	N	
9	EVFXOPEK		N	N	Y	N	N	
9	HLFXHSKK		N	N	Y	N	N	
9	LPGXSF-SIF		N	N	N	N	Y	
9	LPFXHSKKK		N	N	Y	N	N	
10	VL AALAAVXL		N	Y	N	N	N	
10	HLFXHSKKK		N	N	Y	N	N	
10	AAXNMWTRGEM		N	N	Y	N	N	
10	YLLPHTGPHV	L2.LV10	N	Y	N	N	N	1
9	HPGCSF-SIF		N	N	N	N	Y	
9	LPVCSF-SIF		N	N	N	N	Y	
9	LPGCSF-SYF		N	N	N	N	Y	
9	LPFCFS-SIF		N	N	N	N	Y	
9	LPGCSF-SIF		N	N	N	N	Y	
9	LPGCSF-SPF		N	N	N	N	Y	
9	LPGCSF-SIL		N	N	N	N	Y	
9	PPVMHGCPI		N	N	N	N	Y	
10	KPTLLIGPTPI		N	N	N	N	Y	
10	APTLVARMII		N	N	N	N	Y	
9	SFNGSNPSI		N	N	N	N	Y	
10	LPHTGPHLGI		N	N	N	N	Y	
9	SPQARVET		N	N	N	N	Y	
9	LPGCSF-SIL		N	N	N	N	Y	
9	DPTHTSTHIL		N	N	N	N	Y	
10	SPGALVGVVI		N	N	N	N	Y	
10	TPLLYRLGAI		N	N	N	N	Y	
9	TISGVLWQV		N	Y	N	N	N	Nb
9	SISGVLWQV		N	Y	N	N	N	Nb
9	SILMAFTASV		N	Y	N	N	N	Nb
9	GLPDCIMLY		N	Y	N	N	N	Nb
10	KLVALGVNAV		N	Y	N	N	N	Nb
10	YLLPSRGPKL		N	Y	N	N	N	Nb
10	KLSQLGINAV	L1V2.L110	N	Y	N	N	N	Nb
10	YALPRHGRIL		N	Y	N	N	N	Nb
10	VFFNLIGWV		N	N	N	N	N	
10	KLVSLGVNAV		N	Y	N	N	N	
9	CINGYCWTA	I2. V A9	N	Y	N	N	N	
9	CANGVCWTV	IA2. V9	N	Y	N	N	N	

JICV ANALOGS

AA	Sequence	Fixed Nonren.	A1 Mottl	A2 Super Mottl	A3 Super Mottl	A24 Mottl	B7 Super Mottl	1* Anchor Fixer
9	CVNGVCWAV 40		N	Y	N	N	N	

Table XXIII. Immunogenicity of identified supermotif-bearing peptides

Immunogenicity												
Human ^a												
Supermotif	Peptide	Sequence	Protein	Position	Barnaba:		Chisari		Pape	overall	Transgenic mice ^b	
					patients	contacts					Frequency	Response
A2	1073 05	LIFNII.GGWV	NS4	1812	1/6	7/17	2/21	0/6	10/50	6/6	6.4 (1.7)	
	1090 18	FELLADARV	NS1/E2	728	2/6	7/17	1/21	0/6	10/50	5/6	9.5 (3.0)	
	1013 02	YLVAAYQATV	NS4	1590	1/6	4/17	1/21	0/6	6/50	5/6	8.5 (3.7)	
	1090 22	RLIVFPDLGV	NS5	2578	2/6	5/17	0/21	0/6	7/50	0/6	-	
	1013 1002	DLMGYIPLV	Core	132	2/6	7/17	1/21	1/6	11/50	5/6	8.8 (2.6)	
	24 0073	WMNRLIAFA	NS4	1920	1/6	3/17	2/21	1/6	7/50	0/6	-	
	24 0075	VLVGGVLAA	NS4	1666	1/6	6/17	3/21	1/6	11/50	0/6	-	
	1174 08	HMMWNEISGI	NS4	1769	3/6	3/17	2/21	0/6	8/50	6/6	6.4 (1.7)	
	1073 06	ILAGYGAGV	NS4	1851	2/6	3/17	0/21	0/6	5/50	3/6	54.7 (3.3)	
	1073 07	YLLPRRGPRLL	Core	35	2/6	5/17	7/21	1/6	17/50	4/6	59.1 (7.2)	
A3	24 0071	LEFILLADA	NS1/E2	726	2/6	9/17	0/21	0/6	11/50	0/6	-	
	1 0119	YLVTRHADV	NS3	1131	6/6	10/17	0/21	1/6	17/50	0/6	-	
	1 0952	KTSERSQPR	Core	51	2/16	1/4	3/12	0/6	6/38	3/6	23.4 (1.3)	
	1073 11	RLGVRATRK	Core	43	4/16	1/4	7/12	1/6	13/38	3/6	42.2 (1.2)	
	1 0955	QLFTFSPPRR	ENV	290	1/16	0/4	6/12	1/6	8/38			
	1073 13	RMVYVGVEHR	NS1/E2	632	5/16	1/4	4/12	1/6	11/38	2/6	2.8 (1.1)	
	1 0123	LIFCHSKKK	NS3	1396	6/16	1/4	4/12	2/6	13/38	3/6	4.4 (1.1)	
	1073 10	GVAGALVAFK	NS4	1863	3/16	0/4	6/12	2/6	11/38	6/6	56.5 (1.7)	
	24 0090	VAGALVAFK	NS4	1864	4/16	1/4	6/12	0/4	11/38	1/6	7.1	
	24 0086	TLGFGAYMSK	NS3	1262	6/16		2/12	2/5	10/33			
B7	1145 12	LPGCSFSIF	Core	169			2	3/10	5			

Table XXIV. Human and murine MHC-peptide binding assays established using purified MHC molecules and gel filtration chromatography

A. Class I binding assays

Species	Antigen	Allele	Cell line	Radiolabeled peptide		Notes
				Source	Sequence	
Human	A1	A*0101	Steinlin	Hu. J chain 102-110	YTAVVPLVY	no NEN in PI cocktail
	A2	A*0201	JY	HBVc 18-27 F6->Y	FLPSDYFPSV	"
	A2	A*0202	P815 (transfected)	HBVc 18-27 F6->Y	FLPSDYFPSV	"
	A2	A*0203	FUN	HBVc 18-27 F6->Y	FLPSDYFPSV	"
	A2	A*0206	CLA	HBVc 18-27 F6->Y	FLPSDYFPSV	"
	A2	A*0207	721.221 (transfected)	HBVc 18-27 F6->Y	FLPSDYFPSV	"
	A3		GM3107	non-natural (A3CON1)	KVFPYALINK	"
	A11		BVR	non-natural (A3CON1)	KVFPYALINK	"
	A24	A*2402	KAS116	non-natural (A24CON1)	AYIDNNYKFE	"
	A31	A*3101	SPACH	non-natural (A3CON1)	KVFPYALINK	"
	A33	A*3301	LWAGS	non-natural (A3CON1)	KVFPYALINK	"
	A28/68	A*6801	CIR	HBVc 141-151 T7->Y	STLPETYVVR	"
	A28/68	A*6802	AMAI	HBV pol 646-654 C4->A	FTQAGTAL	"
	B7	B*0702	GM3107	A2 signal seq. 5-13 (L7->Y)	APRTLVVLL	"
	B8	B*0801	Steinlin	HBV gp 586-593 Y1->F, Q5->R 60s	FLKDYQL	"
	B27	B*2705	LG2		FRYNGLIHR	"
	B35	B*3501	CIR, BVR	non-natural (B35CON2)	FPEKYAAAF	"
	B35	B*3502	TISI	non-natural (B35CON2)	FPEKYAAAF	"
	B35	B*3503	ELIM	non-natural (B35CON2)	FPEKYAAAF	"
	B44	B*4403	PITOUT	non-natural (B35CON2)	AEMGKYSFY	"
	B51		KAS116	non-natural (B35CON2)	FPEKYAAAF	"
	B53	B*5301	AMAI	non-natural (B35CON2)	FPEKYAAAF	"
	B54	B*5401	KT3	non-natural (B35CON2)	FPEKYAAAF	"
	Cw4	Cw*0401	CIR	non-natural (C4CON1)	QYDDAVYKL	"
	Cw6	Cw*0602	721.221 transfected	non-natural (C6CON1)	YRHIDGCVL	"
	Cw7	Cw*0702	721.221 transfected	non-natural (C6CON1)	YRHIDGCVL	"
Mouse	D ^b		EI.4	Adenovirus E1A P7->Y	SGPSNTYPEI	"
	K ^b		EI.4	VSV NP 52-59	RGYVFQGL	"
	D ^d		P815	HBV-IIIIB ENV G4->Y	RGPYRAFTI	"
	K ^d		P815	non-natural (KdCON1)	KFNPMKTYI	"
	L ^d		P815	HBVs 28-39	IPQSLDSYWTSL	"

Table XXIV. Human and murine MHC-peptide binding assays established using purified MHC molecules and gel filtration chromatography

Species	Antigen	Allele	Cell line	Radiolabeled peptide		Notes
				Source	Sequence	
Human	DR1	DRB1*0101	LG2	HA Y307-319	YPKYVKQNTLKLAT	
	DR2	DRB1*1501	1.466.1	MHP 88-102Y	VVHFFKNIVTPRTPPY	
	DR2	DRB1*1601	1.242.5	non-natural (760.16)	YAAFAAAKTAATAFA	
	DR3	DRB1*0301	MAT	MT 65KD Y3-13	YKTIATFDEARR	
	DR4w4	DRB1*0401	Preiss	non-natural (717.01)	YARFQSOITLKQKT	optimal assay pH is 4.5
	DR4w10	DRB1*0402	YAR	non-natural (717.10)	YARFOROTLKAAA	
	DR4w14	DRB1*0404	BIN 40	non-natural (717.01)	YARFQSOITLKQKT	
	DR4w15	DRB1*0405	K13	non-natural (717.01)	YARFQSOITLKQKT	
	DR7	DRB1*0701	Pitout	Tet. tox. 830-843	QYIKANSKEGITE	
	DR8	DRB1*0802	OIL	Tet. tox. 830-843	QYIKANSKEGITE	
	DR8	DRB1*0803	LUY	Tet. tox. 830-843	QYIKANSKEGITE	
	DR9	DRB1*0901	IID	Tet. tox. 830-843	QYIKANSKEGITE	
	DR11	DRB1*1101	Sweig	Tet. tox. 830-843	QYIKANSKEGITE	
	DR12	DRB1*1201	Hertuf	unknown eluted peptide	EALIHQIKINPYVLS	
	DR13	DRB1*1302	H0301	Tet. tox. 830-843 S>A	QYIKANAKFEGITE	
	DR51	DRB5*0101	GM3107 or 1416.3	Tet. tox. 830-843	QYIKANAKFEGITE	
	DR51	DRB5*0201	1.255.1	HA 307-319	PKYVKQNTLKLAT	
	DR52	DRB3*0101	MAT	Tet. tox. 1272-1284	NGQIGNDPNRDL	
	DR53	DRB4*0101	1.257.6	non-natural (717.01)	YARFQSOITLKQKT	
Mouse	DQ3.1	DQA1*0301/DQB1*0301	PI	non-natural (ROIIV)	YAHAAIAHAIAHAHAA	no NEM in PI mix
	IA ^b		DB27.4	non-natural (ROIIV)	YAHAAIAHAIAHAHAA	optimal assay pH is 5.5
	IA ^d		A20	non-natural (ROIIV)	YAHAAIAHAIAHAHAA	
	IA ^a		CH-12	HEL 46-61	YNTDCESTDYGH QINSR	optimal assay pH is 5.0
	IA ^s		LS102.9	non-natural (ROIIV)	YAHAAIAHAIAHAIAHAA	
	IA ^u		91.7	non-natural (ROIIV)	YAHAAIAHAIAHAIAHAA	
	IE ^d		A20	Lambda repressor 12-26	YLEDARRKKKAIYFKKK	optimal assay pH is 5.0
	IE ^a		CH-12	Lambda repressor 12-26	YLEDARRKKKAIYFKKK	optimal assay pH is 5.0

Table XXV. Monoclonal antibodies used in MHC purification.

Monoclonal antibody	Specificity
W6/32	HLA-class I
B123.2	HLA-B and C
IVD12	HLA-DQ
LB3.1	HLA-DR
M1/42	H-2 class I
28-14-8S	H-2 D ^b and L ^d
34-5-8S	H-2 D ^d
B8-24-3	H-2 K ^b
SF1-1.1.1	H-2 K ^d
Y-3	H-2 K ^b
10.3.6	H-2 IA ^k
14.4.4	H-2 IE ^d , IE ^k
MKD6	H-2 IA ^d
Y3JP	H-2 IA ^b , IA ^s , IA ^u

Table XXVI: HCV-derived conserved high algorithm A*0201-binding peptides

Peptide	Molecule	1st Position	Sequence	Consv.	A2-supertype binding capacity (IC50 nM)						A2 XRN
					A*0201	A*0202	A*0203	A*0206	A*6802		
1073.05	NS4	1812	LLFNILGGWV	85	4.2	113	3.2	19	33	5	
1090.18	NS1/E2	728	FLLADARV	92	18	90	149	247	111	5	
1013.02	NS4	1590	YLVAYQATV	85	20	39	16	82	33	5	
1090.22	NS5	2611	RLVFPDLGV	79	56	391	10	370	8000	4	
1013.1002	CORE	132	DLMGYIPLV	79	80	4778	204	481	12	4	
24.0073	NS4	1920	WMNRLIAFA	100	122	130	3.3	1609	400	4	
24.0075	NS4	1666	VLVGGVLA	85	185	331	32	308	3077	4	
1174.08	NS4	1769	HMWNFISGI	92	15	10750	77	132	7547	3	
1073.06	NS4	1851	ILAGYGAGV	79	116	143	5.0	755	889	3	
1073.07	CORE	35	YLLPRRGPR	92	125	6143	455	416	10256	3	
24.0071	NS1/E2	726	LLFLLADA	100	217	287	455	3364	3077	3	
1.0119	LORF	1131	YLVTRHADV	85	455	2048	3.6	71	3077	3	
24.0065	NS4	1891	ILSPGALVV	92	238	10750	27	1028	3077	2	
1013.12	NS1/E2	686	ALSTGLIHL	85	313	7167	45	18500	10256	2	
939.14	NS1/E2	696	HLHQNIQVDV	85	500	3071	19	1370	10811	2	
1090.21	NS5	2918	RLHGLSAFSL	79	179	782	625	18500	12500	1	

Table XXVII: HCV-derived conserved high algorithm A*03 and/or A*11 binding peptides

Peptide	Molecule	1st Position	Sequence	Consv.	A3-supertype binding capacity (IC50 nM)						
					A*03	A*11	A*3101	A*3301	A*6801	A3 XRN	
1.0952	CORE	51	KTSERSQPR	92	69	94	67	1813	145	4	
1073.11	CORE	43	RLGVRATRK	79	12	207	429	-	-	3	
1.0955	ENV1	290	QLFTSPRR	79	15	182	621	3766	3	3	
1073.13	NS1/E2	632	RMYYVGVEHR	100	15	300	95	9667	1778	3	
1.0123	NS3	1396	LIFCHSKKK	100	20	32	2535	24167	333	3	
1073.10	NS4	1863	GVAGALVAFK	85	28	4	3273	26364	118	3	
24.0090	NS4	1864	VAGALVAFK	85	46	7	3750	11600	258	3	
24.0086	NS3	1262	LGFGAYMSK	85	136	21	2950	22308	222	3	
1174.16	NS1/E2	557	WMNSTGFTK	79	208	74	12857	690	1429	2	
1073.14	NS3	1261	TLGFGAYMSK	85	136	98	-	22308	8889	2	
1090.23	LORF	1183	AVCTRGVAK	79	423	240	16364	-	-	2	
1090.24	NS5	2596	EVFCVQPEK	85	13750	222	-	-	18	2	
24.0103	NS1/E2	647	AACNWTGER	85	36667	429	400	5273	4444	2	
1073.16	NS3	1232	HLHAPTSGK	85	19	2500	-	-	2857	1	
1073.12	NS3	1395	HLIFCHSKKK	100	423	-	20000	-	-	1	
1090.26	NS3	1395	HLIFCHSKK	100	440	10000	-	-	8000	1	

* A dash indicates IC50nM >30,000

Table XXVIII: HCV derived conserved B*0702 binding peptides

A. High conservancy 9- and 10-mer peptides.

Peptide	Molecule	1st Position	Sequence	Consv.	B7-supertype binding capacity (IC50 nM)						B7 XRN
					B*0702	B*3501	B*51	B*5301	B*5401		
1145.12	Core	169	LPGCSFSIF	92	28	90	100	114	6667	4	
15.0048	E2	681	LPALSTGLI	85	157	-	2.8	1500	20000	2	
15.0234	NS3	1620	KPTLHGPTPL	79	3.9	-	27500	-	-	1	
15.0247	NS5	2835	APTLWARMIL	79	6.3	-	5500	-	-	1	
15.0042	CORE	99	SPRGSRPSW	79	14	-	11000	-	-	1	
15.0039	Core	57	QPRGRRQPI	92	24	-	-	-	-	1	
15.0218	Core	37	LPRRGPRLG	92	29	-	6111	-	4000	1	
15.0060	NS5	2615	SPGQRIEFL	79	46	-	27500	-	-	1	
15.0043	Core	111	DPRRRSRNL	85	324	-	-	-	-	1	
15.0063	NS5	2835	APTLWARMIL	79	344	-	4583	-	-	1	
1292.17	NS5	2317	PPVVHGCPL	79	393	-	-	-	-	1	
15.0239	NS4	1893	SPGALVGVV	79	423	-	3438	-	-	1	
15.0235	NS3	1621	TPILYRLGAV	92	458	-	6875	-	909	1	

Table XXVIII: HCV derived conserved B*0702 binding peptides

B. Additional HCV derived B7 supermotif peptides.

Peptide	Molecule	1st Position	Sequence	Consv.	B7-supertype binding capacity (IC50 nM)					
					B*0702	B*3501	B*51	B*5301	B*5401	B7 XRN
29.0035	NS3	1378	IPFYGKAI	92	458	-	46	-	50	3
29.0040	Core	37	LPRRGPR	92	0.85	-	306	-	5000	2
29.0036	Core	137	IPLVGAPL	79	13	2250	79	-	2857	2
16.0187	NS1/E2	680	LPCSFTLPA	64	423	24000	9167	-	15	2
29.0039	Core	169	LPGCSFSI	92	500	200	932	620	6250	2
15.0219	Core	142	APLGGARAL	71	9.5	-	-	-	12500	1
29.0031	NS5	2869	APTLWARM	79	13	-	4583	-	4348	1
15.0231	NS3	1512	RPSGMFDSSV	71	153	-	-	-	-	1
29.0085	NS5	2474	LPIINALNSL	57	220	18000	1170	-	11111	1
29.0037	NS5	2608	KPARLIVF	85	367	-	3235	-	16667	1
15.0237	NS4	1789	NPALASLMAF	71	393	9000	5000	-	-	1
29.0118	NS5	2869	APTLWARMILM	79	423	-	-	-	3030	1
29.0042	NS4	1720	LPYIEQGM	85	423	-	1375	-	7692	1

C. Engineered analogs of B7 supermotif peptides.

Peptide	Molecule	1st Position	Sequence	Consv.	B7-supertype binding capacity (IC50 nM)					
					B*0702	B*3501	B*51	B*5301	B*5401	B7 XRN
1145.12	Core	169	LPGCSFSIF	92	28	90	100	114	6667	4
1292.24	Core	169	LPGCSFSII		37	4364	5.3	262	1056	3
1145.13	Core	169	FPGCSFSIF		19	1.6	132	3.2	6.7	5

* A dash indicates IC50 nM >30,000.

Table XXIX: HCV-derived A1- and A24-motif containing peptides**A. A1-motif peptides**

Peptide	Molecule	Position	Sequence	Conserv.	HLA-A*0101 binding (IC50 nM)
13.0019	NS5	2922	LSAFSLHSY	79	31
1.0509	NS5	2921	GLSAFSLHSY	79	61
1069.62	NS3	1128	CTCGSSDLY	79	68
24.0093	NS5	2129	EVDGVR LHRY	100	167
13.0016	NS3	1241	KSTKVPAAAY	85	1923
1.0125	NS3	1525	CYDAGCAWY	79	4032
24.0008	E1	206	DCSNSSIVY	85	16667
24.0094	NS5	2720	TNSKGQNCGY	100	-
24.0096	NS3	1240	GKSTKVPAAAY	85	-
24.0100	NS3	1292	TGAPITYSTY	85	-
	NS3	1263	VAATLGFGAY	100	
	NS5	2639	VMGSSYGFQY	79	
	NS5	2640	MGSSYGFQY	79	

A dash indicates IC50 nM >25000

B. A24 -motif peptides

Peptide	Molecule	Position	Sequence	Conserv.	HLA-A*2402 binding (IC50 nM)
24.0092	NS4	1765	FWAKHMWNF	85	1.7
13.0075	NS4	1778	QYLAGLSTL	100	250
1073.18	NS1/E2	636	MYVGGVEHRL	92	444
13.0074	NS3	1297	TYSTYGKFL	85	522
13.0134	NS5	2647	QYSPGQ RVEF	79	667
24.0091	NS4	1772	NFISGIQYL	100	706
13.0131	Core	135	GYIPLVGAPL	79	2105
24.0108	Core	173	SFSIFLLALL	100	2927
13.0132	NS3	1248	AYAAQGYKVL	79	13333
13.0133	NS4	1859	GYGAGVAGAL	85	-
1174.08	NS4	1769	HMWNFISGI	93	
	E1	317	RMAWDMMMWNW	85	
	NS1/E2	635	RMVVGVEHRL	93	
	NS3	1422	YYRGLDVSVI	100	
	NS3	1468	DFSLDPTFTI	100	
	NS3	1608	SWDQMWKCL	79	
	NS3	1664	TWVLVGGVL	85	
	NS4	1732	QFKQKALGL	85	
	NS4	1732	QFKQKALGLL	85	
	NS4	1765	FWAKHMWNFI	85	
	NS4	1919	QWMNRLIAF	100	
	NS5	2241	LWRQEMGGNI	85	
	NS5	2669	GFSYDTRCF	79	
	NS5	2875	RMILMTHFF	85	

A dash indicates IC50 nM >25000

Table XXX: Immunogenicity of A2-supertype cross-reactive binders

Immunogenicity										
Human ^a										
Peptide	Sequence	Protein	Position	Barnaba; Barnabat.		Chisari	Pape	overall	Frequency	Response
				patients	contacts					
1073.05	LLFNILGGWV	NS4	1812	1/6	7/17	2/21	0/6	10/50	6/6	6.4 (1.7)
1090.18	FLLADARV	NS1/E2	728	2/6	7/17	1/21	0/6	10/50	5/6	9.5 (3.0)
1013.02	YL.VVAYQATV	NS4	1590	1/6	4/17	1/21	0/6	6/50	5/6	8.5 (3.7)
1090.22	RLIVFPDLGV	NS5	2578	2/6	5/17	0/21	0/6	7/50	0/6	-
1013.1002	DLMGYIPLV	Core	132	2/6	7/17	1/21	1/6	11/50	5/6	8.8 (2.6)
24.0073	WMNRLIAFA	NS4	1920	1/6	3/17	2/21	1/6	7/50	0/6	-
24.0075	VL.VGGVLA	NS4	1666	1/6	6/17	3/21	1/6	11/50	0/6	-
1174.08	HMMWNFISGI	NS4	1769	3/6	3/17	2/21	0/6	8/50	6/6	6.4 (1.7)
1073.06	IL.AGYGAGV	NS4	1851	2/6	3/17	0/21	0/6	5/50	3/6	54.7 (3.3)
1073.07	YL.LPRGRPRL	CORE	35	2/6	5/17	7/21	1/6	17/50	4/6	59.1 (7.2)
24.0071	LLFLLADA	NS1/E2	726	2/6	9/17	0/21	0/6	11/50	0/6	-
1.0119	YL.VTRHADV	NS3	1131	6/6	10/17	0/21	1/6	17/50	0/6	-

a. Data shown represents the number of positive responses over the total number of patients or contacts examined.

b. Frequency represents the number of positive responses over the total number of mice examined. Response indicates the average magnitude (standard deviation) of the response in positive animals, measured in lytic units.

Table XXXI: Immunogenicity of A3-supertype cross-reactive binders

Peptide	Sequence	Protein	Position	Immunogenicity						
				Human ^a					Transgenic mice ^b	
				Barnaba						
				Barnaba patients	Barnaba contacts	Chisari	Pape	overall		Frequency
1.0952	KTSEKQPR	CORE	51	2/16	1/4	3/12	0/6	6/38	3/6	23.4 (1.3)
1073.11	RLGVRATRK	CORE	43	4/16	1/4	7/12	1/6	13/38	3/6	42.2 (1.2)
1.0955	QLITFSPPR	ENV	290	1/16	0/4	6/12	1/6	8/38		
1073.13	RMVVGGEVHIR	NS1/E2	632	5/16	1/4	4/12	1/6	11/38	2/6	2.8 (1.1)
1.0123	LIFCHSKKK	NS3	1396	6/16	1/4	4/12	2/6	13/38	3/6	4.4 (1.1)
1073.10	GVAGALVAFK	NS4	1863	3/16	0/4	6/12	2/6	11/38	6/6	56.5 (1.7)
24.0090	VAGALVAFK	NS4	1864	4/16	1/4	6/12	0/4	11/38	1/6	7.1
24.0086	TLFGAYMSK	NS3	1262	6/16		2/12	2/5	10/33		

a. Data shown represents the number of positive responses over the total number of patients or contacts examined.

b. Frequency represents the number of positive responses over the total number of mice examined. Response indicates the average magnitude (standard deviation) of the response in positive animals, measured in lytic units.

Table XXXII. Candidate HCV-derived HTL epitopes

Selection criteria	Peptide	Sequence	Source	Conservancy	
				Total	Core
A DR-supermotif conserved 15mers	1283.01	GQIVGGVYLLPRRGPR	HCV Core 28	93	93
	1283.02	VYLLPRRGPRLGVRA	HCV Core 34	93	93
	1283.03	GWLLSPRGSRPSWGPT	HCV Core 95	79	79
	1283.04	LGKVIDTLTCGFADL	HCV Core 119	79	86
	1283.05	IDTLTCGFADLMGYI	HCV Core 123	86	86
	1283.06	ADLMGYIPLVGAPLG	HCV Core 131	79	79
	1283.07	GVRVLEDGVNYATGN	HCV Core 154	86	86
	1283.08	GVNYATGNLPGCSFS	HCV Core 161	79	86
	1283.09	GCSFSIFLLALLSCL	HCV Core 171	86	100
	1283.10	GHRMAWDMMMNWSPPT	HCV E1 315	86	86
	1283.11	CGPVYCFTPSPVVVG	HCV NS1/E2 506	93	93
	1283.12	VYCFTSPVVVGTTD	HCV NS1/E2 509	93	93
	1283.13	GNWFGCTWMNSTGFT	HCV NS1/E2 550	79	86
	1283.14	FTTLPALSTGLIHLH	HCV NS1/E2 684	79	86
	1283.17	DLYLVRHADVIPVR	HCV NS3 1134	79	79
	1283.18	RAAVCTRGVAKAVDF	HCV NS3 1186	79	79
	1283.20	AQGYKVLVLNPSVAA	HCV NS3 1251	79	100
	1283.21	GKVLVLNPSVAATL	HCV NS3 1253	100	100
	1283.22	VLVLNPSVAATLGFG	HCV NS3 1256	100	100
	1283.23	GTVLDQAETAGARLV	HCV NS3 1335	86	86
	1283.24	GARLVVLATATPPGS	HCV NS3 1345	79	86
	1283.25	GRHLIFCHSKKKCDE	HCV NS3 1393	100	100
	1283.27	DSVIDCNTCVTQTVD	HCV NS3 1454	86	86
	1283.28	TVDFSLDPTFTIETT	HCV NS3 1466	79	100
	1283.30	FTGLTHIDAHFLSQT	HCV NS3 1567	93	93
	1283.31	YLVAQATVCARAQA	HCV NS3 1591	79	93
	1283.32	KPTLHGPTPLLYRLG	HCV NS4 1620	79	79
	1283.33	LEVVTSTWVLVGGVL	HCV NS4 1658	86	86
	1283.34	TWVLVGGVLAALAAAY	HCV NS4 1664	86	86
	1283.35	AEQFKQKALGLLQTA	HCV NS4 1730	86	86
	1283.40	PAILSPGALVVGVCVA	HCV NS4 1889	79	93
	1283.41	GALVVGVVCAAILRR	HCV NS4 1895	79	79
	1283.42	CAAILRRHVGPGEGA	HCV NS4 1903	79	79
	1283.43	AVQWMNRLIAFASRG	HCV NS4 1917	100	100
	1283.44	MNRLIAFASRGNHVS	HCV NS4 1921	86	100
	1283.48	ANLLWRQEMGGNITR	HCV NS5 2238	86	86
	1283.49	RQEMGGNITRVESEN	HCV NS5 2243	86	86
	1283.52	ARLIVFPDLGVRVCE	HCV NS5 2610	79	79
	1283.53	FPDLGVRVCEKMALY	HCV NS5 2615	79	100
	1283.54	GVRVCEKMALYDVVS	HCV NS5 2619	79	100
	1283.56	QPEYDLELITSCSSN	HCV NS5 2808	79	93
	1283.57	LELITSCSSNVSAH	HCV NS5 2813	79	100
	1283.58	PTLWARMILMTHFFS	HCV NS5 2870	79	86
	1283.59	LHGLSAFSLHSYSPG	HCV NS5 2919	79	79
	1283.60	AFSLHSYSPGEINRV	HCV NS5 2924	79	79

Table XXXII. Candidate HCV-derived HTL epitopes

Selection criteria	Peptide	Sequence	Source	Conservancy	
				Total	Core
B High algorithm conserved core	1283.15	VVLLFLLADARVCS	HCV NS1/E2 724	29	100
	1283.16	SKGWRLLAPITAYAQ	HCV NS3 1025	29	79
	1283.19	PQTFQVAHLHAPTGS	HCV NS3 1225	43	85
	1283.26	DVVVVATDALMTGYT	HCV NS3 1436	43	79
	1283.29	WESVFTGLTHIDAHF	HCV NS3 1563	43	92
	1283.45	LTSMLTDPShITAET	HCV NS5 2176	57	100
	1283.46	ASQLSAPSLKATCTT	HCV NS5 2208	50	79
	1283.47	DADLIEANLLWRQEM	HCV NS5 2232	50	85
	1283.50	SYTWTGALITPCAAE	HCV NS5 2456	64	79
	1283.51	TTIMAKNEVFCVQPE	HCV NS5 2589	64	85
	1283.55	GSSYGFQYSPGQORVE	HCV NS5 2641	71	79
	1283.61	ASCLRKLGVPLRVW	HCV NS5 2939	50	85
C Collaborator	F098.03	AAAYAAQGYKVLVLNPSVAAT	HCV NS3 1242-1261	71	100
	F098.04	GYKVLVLNPSVAATLGFGAY	HCV NS3 1248-1267	100	
	F098.05	GYKVLVLNPSVAAT	HCV NS3 1248-1261	100	
	F134.01	RRPQDVKFPGGGQIVGGVY	HCV Core 17-35	86	
	F134.02	DKKFPGGGQIVGGVYLLPRR	HCV Core 21-40	86	
	F134.03	GYKVLVLNPSVAATLGFGAY	HCV NS3 1253-1272	100	
	F134.04	TLHGPTPLLYRLGAVQNEIT	HCV NS4 1622-1641		79
	F134.05	NFISGIQYLAGLSTLPGNPA	HCV NS4 1772-1791	100	
	F134.06	LLFNILGGWVAAQLAAPGAA	HCV NS4 1812-1831		86
	F134.07	GPGEAVQWMNRLIAFASRG	HCV NS4 1912-1931	86	100
	F134.08	GEGAVQWMNRLIAFASRGNHV	HCV NS4 1914-1934	100	
	Pape 21	AIPLEVIKGGRLIFCHSKR	HCV NS3 1379-1398	21	100
	Pape 22	GRHLIFCHSKRKCDLATKL	HCV NS3 1388-1407		100
	Pape 29	SVIDCNTCVTQTVDFSLDPT	HCV NS3 1450-1469	86	
D DR3 motif	35.0102	GVRVLEDGVNYATGN	HCV 154	86	86
	35.0103	SAMYVGDLCSVFLV	HCV 273	57	86
	35.0104	GHRMAWDMMMNWSPT	HCV 315	86	86
	35.0105	SDLYLVTRHADVIPV	HCV 1133	79	86
	35.0106	VVVVATDALMTGYTG	HCV 1437	42	86
	35.0107	TVDFSLDPTFTIETT	HCV 1466	79	100
	35.0108	DSSVLCECYDAGCAW	HCV 1518	71	93
	35.0109	GLPVCQDHLEFVESV	HCV 1552	42	86
	35.0110	GMQLAEQFKQKALGL	HCV 1726	57	86
	35.0111	PTHYVPESDAAARVT	HCV 1936	86	86
	35.0112	GSQLPCEPEPDVAVL	HCV 2162	64	86
	35.0113	LTSMLTDPShITAET	HCV 2176	57	100
	35.0114	MPPLEGEPGDPDLSD	HCV 2401	79	100
	35.0115	QPEYDLELITSCSSN	HCV 2808	79	93
	1283.25	GRHLIFCHSKKKCDE	HCV NS3 1393-1407		

Table XXXIII. HLA-DR screening panels

Screening Panel	Antigen	Alleles	Representative Assay			Phenotypic Frequencies						
			Allele	Alias	Cauc.	Blk.	Jpn.	Chn.	Hisp.	Avg.		
Primary	DR1	DRB1*0101-03	DRB1*0101	(DR1)	18.5	8.4	10.7	4.5	10.1	10.4		
	DR4	DRB1*0401-12	DRB1*0401	(DR4w4)	23.6	6.1	40.4	21.9	29.8	24.4		
	DR7	DRB1*0701-02	DRB1*0701	(DR7)	26.2	11.1	1.0	15.0	16.6	14.0		
	Panel total				59.6	24.5	49.3	38.7	51.1	44.6		
Secondary	DR2	DRB1*1501-03	DRB1*1501	(DR2w2.61)	19.9	14.8	30.9	22.0	15.0	20.5		
	DR2	DRB5*0101	DRB5*0101	(DR2w2.62)	-	-	-	-	-	-		
	DR9	DRB1*0901.09012	DRB1*0901	(DR9)	3.6	4.7	24.5	19.9	6.7	11.9		
	DR13	DRB1*1301-06	DRB1*1302	(DR6w19)	21.7	16.5	14.6	12.2	10.5	15.1		
	Panel total				42.0	33.9	61.0	48.9	30.5	43.2		
Tertiary	DR4	DRB1*0405	DRB1*0405	(DR4w15)	-	-	-	-	-	-		
	DR8	DRB1*0801-5	DRB1*0802	(DR8w2)	5.5	10.9	25.0	10.7	23.3	15.1		
	DR11	DRB1*1101-05	DRB1*1101	(DR5w11)	17.0	18.0	4.9	19.4	18.1	15.5		
	Panel total				22.0	27.8	29.2	29.0	39.0	29.4		
Quaternary	DR3	DRB1*0301-2	DRB1*0301	(DR3w17)	17.7	19.5	0.4	7.3	14.4	11.9		
	DR12	DRB1*1201-02	DRB1*1201	(DR5w12)	2.8	5.5	13.1	17.6	5.7	8.9		
	Panel total				20.2	24.4	13.5	24.2	19.7	20.4		

Table XXXIV. HLA-DR binding capacity of target derived peptides: DR-supermotif and algorithm positive peptides.

Peptide	Sequence	Source	Binding capacity (IC50 nM)										DR allele bound	
			DR1	DR2w2B1	DR2w2B2	DR4w4	DR4w15	DR5w11	DR6w19	DR7	DR8w2	DR9		IAb
1283 21	AAVAAQGYK.VI.VI.NPSVAATL	HCV NS3 1253	45	350	-	52	567	143	51	89	288	54	175	9
1283 20	AQGYK.VI.VI.NPSVAA	HCV NS3 1251	6.0	650	-	79	224	74	59	833	175	375	298	9
1283 03	AAVAAQGYK.VI.VI.NPSVAAT	HCV NS3 1242	29	48	483	18	1234	103	11	96	60	240		9
1283 05	GYK.VI.VI.NPSVAAT	HCV NS3 1248-1261	14	39	3695	78	141	75	35	126	21	266		9
1283 04	GYK.VI.VI.NPSVAATL.GF.GAY	HCV NS3 1248-1267	35	42	8154	97	1500	240	41	23	80	20		8
	GF.GAVQWMNRLLAFASRGNIHS	HCV NS4 1914-1935												
1283 44	MINRLAFASRGNIHS	HCV NS4 1921	66	48	1538	6329	585	45	73	227	102	313	147	8
1283 08	GF.GAVQWMNRLLAFASRGNIHV	HCV NS4 1914	32		182	361		345		221	158	6818		6
1283 16	SKGWRLAPITAYAQ	HCV NS3 1025	0.36	125	23	24	152	48	-	962	54	1190	384	8
1283 55	GSSYGfQYSPGQRVE	HCV NS5 2641	11	-	667	417	745	20000	19	156	-	68	571	7
1283 61	ASCLRLGVPLRVW	HCV NS5 2939	5.0	16	217	6250	78	645	2500	862	671	8621	-	7
1283 05	NFISGQYLAGLSTLPGNPA	HCV NS4 1772	10		606	84		29		-	70	441		6

Shading indicates IC50 > 1 nM

Shading indicates IC50 > 1 μ M

A dash (-) indicates IC50 > 20 μ M

Table XXXV. HLA-DR binding capacity of 3 DR3 motif-containing peptides

Peptide	Sequence	Source	DR3 binding (IC50 nM)
35.0106	VVVVATDALMTGYTG	HCV 1437	427
35.0107	TVDFSLDPTTETT	HCV 1466	235
1283.25	GRHLIFCHSKKKCDE	HCV NS3 1393	ND

Table XXXVIa: HCV-derived CTL epitope candidates

Peptide	Molecule	1st Position	Sequence	Consv.	Selection criteria
1073.05	NS4	1812	LLFNILGGWV	85	A2-supertype
1090.18	NS1/E2	728	FLLADARV	92	A2-supertype
1013.02	NS4	1590	YLVAYQATV	85	A2-supertype
1090.22	NS5	2611	RLIVFPDLGV	79	A2-supertype
1013.1002	CORE	132	DLMGYIPLV	79	A2-supertype
24.0073	NS4	1920	WMNRLIAFA	100	A2-supertype
24.0075	NS4	1666	VLVGVLAA	85	A2-supertype
1174.08	NS4	1769	HMWNFISGI	92	A2-supertype
1073.06	NS4	1851	ILAGYGAGV	79	A2-supertype
1073.07	CORE	35	YLLPRRGPR	92	A2-supertype
24.0071	NS1/E2	726	LELLADA	100	A2-supertype
1.0119	LORF	1131	YLVTRHADV	85	A2-supertype
1.0952	CORE	51	KTSERSQPR	92	A3-supertype
1073.11	CORE	43	RLGVRATRK	79	A3-supertype
1.0955	ENV1	290	QLFTFSPPR	79	A3-supertype
1073.13	NS1/E2	632	RMVYVGVEHR	100	A3-supertype
1.0123	NS3	1396	LIFCHSKKK	100	A3-supertype
1073.10	NS4	1863	GVAGALVAFK	85	A3-supertype
24.0090	NS4	1864	VAGALVAFK	85	A3-supertype
24.0086	NS3	1262	TLGFGAYMSK	85	A3-supertype
F104.01	NS5	3003	VGIYLLPNR	79	A31
1145.12	CORE	169	LPGCSFSIF	92	B7-supertype
29.0035	NS3	1378	IPFYGKAI	92	B7-supertype
13.0019	NS5	2922	LSAFSLHSY	79	A1
1069.62	NS3	1128	CTCGSSDIY	79	A1
24.0092	NS4	1765	FWAKHIMWNF	85	A24

Table XXXVIIb: HCV-derived HTL epitope candidates

Region	Peptide	Motif ¹	Sequence
HCV NS3 1025-1039	1283.16	DR	SKGWRLAPITAVAQ
HCV NS3 1242-1267	F98.03	DR	AAVYAAQGYKVLVLPNSVAAT
HCV NS3 1393-1407	1283.25	DR3	GRHLIFCHSKKKKCODE
HCV NS3 1437-1451	35.0106	DR3	VVVVAITDALMTGYTG
HCV NS3 1466-1480	35.0107	DR3	IVDESLDPITFIEIT
HCV NS4 1772-1790	F134.05	DR	NFISGIQYL.AGL.STLPGNPA
HCV NS4 1914-1935	F134.08	DR	GEGAVVQWMNRLIAFASRGNIHV
HCV NS5 2641-2655	1283.55	DR	GSSYGFOYSPGQRYE
HCV NS5 2939-2953	1283.61	DR	ASCLRKLGVPPLRVW

1. Peptides identified on the basis of either the DR P1-P6 supermotif or by use of the DRI-4-7 algorithms are indicated by 'DR'. Peptides identified using the DR3 motif are indicated by 'DR3'.